

A Survey of Perception-Based Visualization Studies by Task

Ghulam Jilani Quadri and Paul Rosen

Abstract—Knowledge of human perception has long been incorporated into visualizations to enhance their quality and effectiveness. The last decade, in particular, has shown an increase in perception-based visualization research studies. With all of this recent progress, the visualization community lacks a comprehensive guide to contextualize their results. In this report, we provide a systematic and comprehensive review of research studies on perception related to visualization. This survey reviews perception-focused visualization studies since 1980 and summarizes their research developments focusing on low-level tasks, further breaking techniques down by visual encoding and visualization type. In particular, we focus on how perception is used to evaluate the effectiveness of visualizations, to help readers understand and apply the principles of perception of their visualization designs through a task-optimized approach. We concluded our report with a summary of the weaknesses and open research questions in the area.

Index Terms—Visualization, perception, graphical perception, visual analytics tasks, evaluation, survey.

1 INTRODUCTION

VISUALIZATION provides valuable assistance in data analysis and decision-making tasks. The human perceptual and cognitive systems are essential in the process of visualization, influencing visual analysis activities, e.g., data exploration, data gathering, and data manipulation. As an example, data exploration requires forming high-level analysis goals, planning actions, and evaluating results effectively, all of which are cognitive activities. Before higher-level cognitive processes analyze data, visualization passes through the human perceptual system impacting the visualization’s utility [1]. The design of the visualization should make it as easy and unambiguous as possible to understand the data. Ultimately, a better understanding of human perception aids visualization design in both a quantitative and qualitative manner [2].

Numerous fields of science have studied perception, including perceptual psychology, visualization, and human-computer interaction. Much of our understanding of perception in visualization is rooted in an early work that ranked the order of *visual encodings* based on their effectiveness for visual judgment [3]. The findings were pivotal in nature and a milestone in research, demonstrating that the application and understanding of perception lead to guidelines for an effective and expressive visual design. Furthermore, numerous works on improving and evaluating visualization’s effectiveness have utilized the knowledge of attention, psychophysics, stimulus, judgment estimation, and perceptual laws to confirm the importance of perception throughout the process of generating a visualization [4], [5], [6], [7], [8], [9], and we have observed a notable increase in

the interest in perception-based visualization studies over recent decades (see Fig. 1).

In this paper, we review perception-focused visualization studies since 1980 and summarize their research developments. Our focus is primarily on information visualization—areas such as scientific visualization, 3D perception, etc., have been surveyed elsewhere, e.g., [10]. Within this context, we create a practical taxonomy of prior studies based upon Amar et al.’s low-level task taxonomy [11], further breaking techniques down by visual encoding and visualization type. In particular, we focus on how perception is used to evaluate the quality of visualizations, to help readers understand and apply the principles of perception to their visualization designs through a task-optimized approach.

We focus our survey on studies that measured the efficacy of the visual design using *graphical perception* when the user is performing various low-level tasks. Furthermore, since perception and cognition are not entirely separable, we discuss some perceptual effects on cognitive performance, e.g., completion time, accuracy, and error rate. Unless otherwise noted, throughout the remainder of this survey, we will use the term *perception* to refer to *graphical perception*.

One of the key challenges with many graphical perception studies is their limited scope and reproducibility [12], caused in part by the difficulty of constructing human studies.

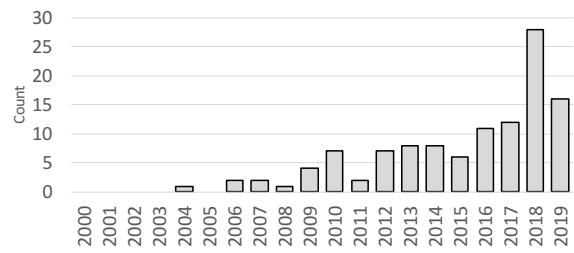


Fig. 1: Publications per year in our survey considering perception and visualization since 2000.

• G.J. Quadri was with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL, 33647.

E-mail: ghulamjilani@usf.edu

• P. Rosen was with the Department of Computer Science and Engineering, University of South Florida, Tampa, FL, 33647.

E-mail: prosen@usf.edu

Human studies are often resource-limited by the types and sizes of data, variety of visualizations or visual encodings, tasks being performed, and the size and diversity of subject pools, to name a few. This all puts transferability of results on tenuous footing. There is mounting evidence of the perceptual efficacy for many tasks and visualization types, but our survey provides a window into open research questions in many other situations. One of our survey's goals is collecting and organizing findings in such a way that the reader can make a judgment of the applicability of results to their context.

1.1 Audience Guide for this Report

In this report, we target students, practitioners, and researchers, each of which will find value in different sections. First, Sect. 2 establishes the structure of our taxonomy, including a discussion of tasks, as well as secondary issues, including the visual encoding and visualization types. Next, Sect. 3 is the main survey of papers, divided by task, which provides background and context to aspects of perception that have been studied in visualization. Sect. 4 concludes the work by discussing key points, limitations, and open research questions. Finally, as an appendix, Sect. 5 provides a brief introduction to some fundamentals of perception for individuals looking to gain basic knowledge of the perceptual principles studied on visualization designs.

An interactive list of the papers we have identified in our survey is available at <<https://usfdatavisualization.github.io/VisPerceptionSurvey/>>.

1.2 Related Works

Optimizing visualization design is a perennial topic.

Design Recommendations. Design choices and recommendations form a critical element of effective visualization. In his book [2], Colin Ware talked about human perception in the context of information visualization design. He aimed at broadly summarizing the design implications of research in perception and suggested explicit design guidelines. In their survey, Healey and Enns focused more specifically on the role that attention and visual memory play in the perception of visualizations [7]. The work highlights how what we see impacts the viewer's accuracy in information judgment. Finally, VisGuides <<https://visguides.org/>> is a web-based forum, which was established to collect practical knowledge of visualization guidelines and feedback on designs [14].

Frameworks. In addition to general guidelines, several researchers have focused on developing robust frameworks for optimizing design. Rensink's framework for reasoning

about perceptions of visualization designs suggests using techniques from vision science [15]. The extended-vision theory asserts that a viewer and visualization system is a single system, whereas the optimal-reduction thesis postulates an optimal visualization. The work focuses on a few of the fundamental questions, e.g.: *What is the best way to measure how a given visualization works? Or, could we determine if its design is optimal?* Recent work by Elliott et al. introduced a design space of experimental methods for empirically investigating the perceptual processes involved in viewing data visualizations to inform visualization design guidelines [16]. The paper provides shared design space and lexicon for facilitating empirical visualization research. Researchers can use this design space to create innovative studies and progress scientific understanding of design choices and evaluations in visualization. In contrast, our paper provides an overview of perception-based studies by surveying papers that have evaluated the effectiveness of various visualization under a variety of tasks.

Evaluation. When optimizing a design, measurements of effectiveness are critical to understanding their impact. Studies often rely upon subject evaluations of visualizations. One important innovation is the introduction of crowdsourcing environments which are faster, less costly, and provide more diverse subject pools than lab-based studies (Fig. 2). Borgo et al. provided a detailed review of the use of crowdsourcing for evaluation in visualization research [13]. In addition to subject evaluations, there are many methods and metrics for quantitatively evaluating the effectiveness of a visualization. Behrisch et al. [17] gave an extensive audit of the state-of-art in quality metrics for various visualization techniques, along with details on a variety of implementation possibilities. The papers we discuss throughout this survey use a combination of both subject evaluations and quantitative measures to formulate their conclusions.

1.3 Systematic Survey Literature

Perception, being a critical part of visualization, has a wide range of applications and is included throughout different visualization-related journals and conferences. In addition, perception and its incorporation to visualization are derived from psychology-related journals, such as the *Journal of Vision, Attention, Perception, and Psychophysics*, and *Psychonomic*. With the survey's objective and the vast availability of perception-based papers, it was challenging to perform a comprehensive literature search. Our taxonomy discusses the application of perception to visualization. Therefore, we focused on visualization journals and conferences.

We identified papers from major visualization journals and conferences between 1980-2019 (see Table 1). We targeted the *ACM*, *IEEE*, and *EG/CGF* libraries to collect the papers using a combination of keywords, including: *perception*, *visualiz/sation*, *evaluation*, *design*, *modeling*, *visual perception*, *attention*, *visual task*, *user study*, *graphical encoding*, and *effectiveness*. The "others" categories in Table 1 are highly cited, pivotal works that were discovered during our search of the primary sources. Using the PRISMA framework (see <<http://www.prisma-statement.org/>>) as a guide, we coded the scope of the survey into categories to screen the literature. We filtered the paper based on the study's objective and

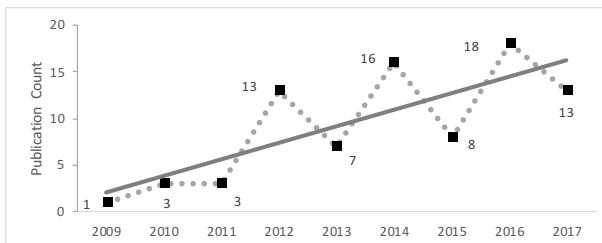


Fig. 2: Reproduction of a plot of the number of publications utilizing crowdsourcing experiments in recent years [13].

TABLE 1: The number of surveyed papers by source.

Sources	Paper Count
IEEE Trans. on Visualization and Computer Graphics (TVCG) IEEE Information Visualization (InfoVis) IEEE Visual Analytics in Science and Technology (VAST) IEEE Pacific Visualization Symposium (PacificVis)	64
ACM Conf. on Human Factors in Information Systems (CHI) including Extended Abstracts ACM Transaction of Graphics (TOG)	30
Computer Graphics Forum (CGF) Eurographics (EG) EG/IEEE VGTC Conference on Visualization (EuroVis) including Short Papers	17
Others—Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV); Journal of Vision; Perception and Psychophysics; Science; Journal of the American Statistical Association; International Conference on Theory and Application of Diagram; Cartographics; Journal of Man-Machine Studies; Behaviour and Information Technology; Others	11

characteristics. We could not include all filtering categories due to space limitations, but some examples are below.

Examples for Inclusion.

- Experiments focused on graphical perception on visual tasks, e.g., [18], or visual design, e.g., [9].
- Experiments focused on the graphical perception of visualization methods, e.g., [19].
- Studies including discussion and suggestions of design guidelines, e.g., [20].
- Experiments on modeling the visualization to improve inference-making, decision-making, or judgment estimation, based on visual channel and graphical perception, e.g., [21], [22].

Examples for Exclusion.

- User-study comparing two or more models, e.g., [23].
- Empirical studies to check on quality metrics of a system by user study, e.g., [24].
- Evaluations of the performance of visualization tools or systems, e.g., [25].
- Empirical studies on graphics or user interfaces, not the visualization, e.g., [26].

In addition to the interactive taxonomy at <https://usfdatalab.github.io/VisPerceptionSurvey/>, you will find (1) a spreadsheet of surveyed papers, (2) the systematic flow of paper collection and filtering, i.e., PRISMA, (3) a paper categorization template, and (4) a summary table with the count of studies.

2 STRUCTURE OF THE TAXONOMY

The method of visually encoding data is usually thought to be the main component of visualization. However, the analysis task is equally, if not more, important. Several evaluation studies have suggested visualization effectiveness is task-dependent [19], and a large body of research seeks to determine which data representations are perceptually optimal for specific low-level tasks, e.g., [2], [3], [20], [27]. Seeing that the vast majority of perceptual studies in visualization had a specific low-level task as the main study objective or a low-level task was used in the evaluation, we centered on that as the main category of the taxonomy. Furthermore,

most papers consider the low-level tasks in the context of a limited subset of visual encodings and/or visualization types. Therefore, each of our low-level task discussions is further split by types of **visual encoding** and **visualization**.

2.1 Low-Level Tasks

We considered two existing task taxonomies as a framework for our survey. We first considered Brehmer and Munzner's taxonomy of abstracted tasks, which is a higher-level taxonomy, i.e., perceptual + cognitive [28]. Since our focus was perception, it did not fit well. Despite perception and cognition not being entirely separable, an ideal task framework, would ensure as little cognition as possible occurs within the task. Ultimately, Amar et al.'s low-level task taxonomy [11] fit better, as the tasks they define require less reasoning about the data. From there, we took its ten low-level tasks: *retrieve value*, *filter*, *compute a derived value*, *find extremum*, *sort*, *determine range*, *characterize distribution*, *find anomalies*, *cluster*, and *correlate*, and we included one derived task, *compare*, for a total of 11 low-level task groups. Each paper was placed into one or more of these task groups. We summarize the tasks using descriptions from prior work:

-  **Retrieve Value** (Sect. 3.1) – “Given a set of specific cases, find attributes of those cases.” [11]
-  **Filter** (Sect. 3.2) – “Given some concrete conditions on attribute values, find data cases satisfying those conditions.” [11]
-  **Compute Derived Value** (Sect. 3.3) – “Given a set of data cases, compute an aggregate numeric representation of those data cases.” [11]
-  **Find Extremum** (Sect. 3.4) – “Find data cases possessing an extreme value of an attribute over its range within the data set.” [11]
-  **Sort** (Sect. 3.6) – “Given a set of data cases, rank them according to some ordinal metric.” [11]
-  **Determine Range** (Sect. 3.5) – “Given a set of data cases and an attribute of interest, find the span of values within the set.” [11]
-  **Find Anomalies** (Sect. 3.7) – “Identify any anomalies within a given set of data cases concerning a given relationship or expectation, e.g., statistical outliers.” [11]
-  **Characterize Distribution** (Sect. 3.8) – “Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set.” [11]
-  **Cluster** (Sect. 3.9) – “Given a set of data cases, find clusters of similar attribute values.” [11]
-  **Correlate** (Sect. 3.10) – “Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.” [11]
-  **Compare** (Sect. 3.11) – “Given a set of data cases, compare any attributes within and between relations of the given set of data cases for a given relationship condition.” [11]

2.2 Visual Encoding

Visual encodings are properties used to encode data in a visualization, including *position*, *length*, *angle*, *area*, *volume*,

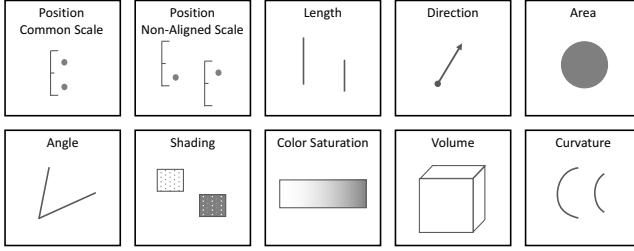


Fig. 3: Reproduction of visual encodings types introduced in Cleveland and McGill's work [3].

shading, direction, curvature, and color (see Fig. 3). The terms graphical encoding, visual channel, visual encodings, and visual properties are often interchanged, but generally, they mean the same thing. Throughout the remainder of this paper, we refer to them as visual encodings.

Effectiveness of Visual Encodings. Understanding the role of perception in the choice of visual encodings is critical to visualization designers. Cleveland and McGill's study evaluated the efficacy of visual encodings by measuring the perceptual magnitude of judgments to determine their accuracy. Mackinlay produced the first comprehensive ranking of visual encodings by data type, as shown in Fig. 4 [29]. The ranking has been further validated and elucidated through numerous follow-up studies, e.g., [9], [18], [20], [30], [31], [32], [33], [34], [35], [36], many of which are further discussed throughout our taxonomy.

Taxonomy on Visual Encodings. Early works in visual encoding defined numerous individual visual channels, e.g., 10 in [3] and 13 in [29]. Given our taxonomy already defines 11 task categories, enumerating all visual encoding types would have resulted in far too many categories. Instead, we combined visual encodings into roughly two main categories: spatial encodings and color encodings, each with two subcategories. Spatial encodings encompass visual encodings having to do with position, size (i.e., length, area, and volume), direction, and shape. Mackinlay, as well as others, have shown that spatial encodings are particularly effective for quantitative data [29]. Color is another important visual encoding that includes properties of color such as hue, saturation, luminance, and opacity. Stimulus-based psychophysics experiments have demonstrated that color can be used to represent ordering [37], category [38], quantity [39], and *perceived difference judgment* [40] in visualization. The

Spatial Position and Shape encodings are combined because of their geometric relationship. The category encompasses any encoding concerning the layout, e.g., position, or shape, e.g., direction/angle or curvature.

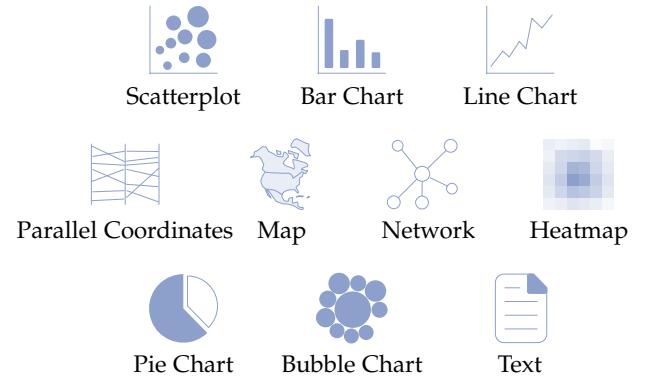
Spatial Size encodings are those where the size of objects, i.e., length, area, or volume, encode the relevant data in the visualization.

Color Hue deals primarily with variations in color (in the colloquial sense), usually dealing with categorical colors or colormaps.

Color Intensity encodings capture data by the intensity of the color representation, e.g., in color saturation, luminance, shading, or opacity, which have combined due to their close inter-relatedness, e.g., changes in opacity, luminance, or saturation can all influence the intensity of a color.

2.3 Visualization

The final category of our taxonomy is visualization type. Most papers studied standard visualization types, quite often with a variation in their design. We extracted all visualization types and combined related types into the following 10 categories and one "other" category:



3 SURVEY

One of our survey's goals was collecting and framing studies in a way that one could generate effective visual designs considering graphical perception. However, our investigation led us to experiments with limited scope, e.g., focus on a small set of visual encodings, visualization types, limited tasks, or limited data set size and variety. Furthermore, most findings are never replicated or externally validated beyond peer review, with only a few exceptions.

These limitations have two effects. First, when the scope of work is limited, it puts transferability of findings on unsteady ground. E.g., if the real-world conditions associated with a data visualization do not match a contrived scenario, often necessary for studying graphical perception, will the findings be relevant? The second limitation is that finding common links between various studies is quite often difficult. E.g., if two studies using different sized data and slightly different tasks have contradictory results, relating the findings in an actionable way is difficult at best. As such, our summary of works in graphical perception tries to highlight these differences while leaving room for the reader to interpret real-world consequences for their particular situation.

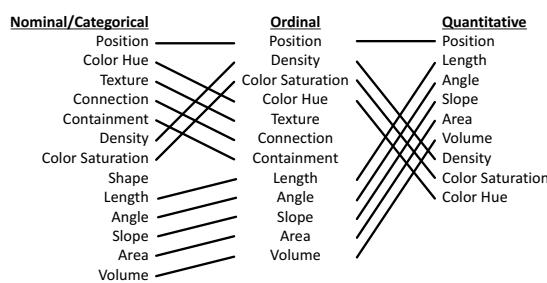


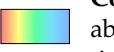
Fig. 4: Reproduction of the Mackinlay visual encoding rankings [29].

3.1 Retrieve Value

 The retrieve value task requires identifying the data or attributes that satisfy a given set of specific criteria. In the literature, retrieve value tasks are often combined with other tasks, e.g., computing a derived value, comparison, sort, etc. We did identify several works that studied the retrieve value task individually with various visualization methods.

3.1.1 Visual Encoding

 **Spatial Position, Shape, & Size.** Visual factors, such as position along common and unaligned scales, have been observed to produce more accurate judgments than length, direction, and angle on simple tasks [3], [29] (see Fig. 4 for the ranking of other encodings). However, a recent study demonstrated in position-based visualizations, e.g., scatterplots, encoding additional attributes with size required more time for retrieving values than alternatively encoding with color [40]. On the other hand, subjects' accuracy improved with size. The findings further explored the symmetry between the size and color as the visual encoding for quantitative information, and they suggested that marks with varied sizes might interfere with decoding the quantity value in position channels. The results further indicated that chart orientation influenced performance by comparing x- and y-faceted charts, with y-faceted charts performing 0.9 times faster. Nevertheless, our understanding of how variations in the visualization design potentially influence user performance on this task is evolving.

 **Color Hue.** Several studies have evaluated people's ability to identify values using color-coded visualizations. In one instance, the performance in identifying or determining value in visualization was shown to be affected by color and other visual features, such as motion or layout [5]. We have identified several studies focused on how color can be applied such that finding targets or reading a value becomes less strenuous. For example, the issue of color vision deficiency, i.e., colorblindness, was studied in automated colormap design [41]. In the study, colormaps were created starting with a single seed color that was used to generate color ramps that mimic designer practices. Their experiments measured the viewer's potential to accurately identify or read a value in the given color-coded scatterplot, heatmap, or choropleth maps. Ultimately, the performance of their automated system was as good as human-designer-specified color ramps [42]. Reasonable precaution is necessary to improve accessibility, i.e., for colorblind, low vision, or other vision impairments, to make the task less challenging and provide an unbiased user response.

 **Color Intensity.** One study showed that lightness in the data point symbol could be an impacting factor for visual tasks, such as locating and identifying values in sparse scatterplots on a white background [43]. This experiment identified how lightness could be modeled as a combination of two opposite power functions and used to determine discriminability. Related to lightness, opacity has been used as a constructive factor in a scatterplot design to mitigate overdraw in tasks, such as to read and identify values for higher-level tasks [44].

3.1.2 Visualizations

A recent study by Saket et al. [18] comparing five visualization types—tabular visualization, scatterplots, bar charts, line charts, and pie chart—evaluated the efficacy of these visualizations with *small data* for all ten of Amar et al.'s tasks [11], in terms of accuracy, time, and user preference (see Fig. 5). For the retrieve value task, they found tabular visualization outperformed all others for accuracy, time, and preference, bar and pie charts performed well in terms of accuracy and time, and scatterplots performed well only in terms of accuracy. A similar comparative study of three multidimensional visualizations—parallel coordinates, scatterplot matrix, and tabular visualization—studied their user performance on analytical and decision-making tasks. For the retrieve value task, parallel coordinates had the highest accuracy. The evaluation also demonstrated that tabular visualization was familiar, accurate, and time-efficient for the retrieve value task [45]. In both studies, tabular visualizations allowed subjects to reach decisions faster with better accuracy levels than the other visualizations because of their familiarity with tables.

 **Scatterplot.** A scatterplot is one of the most effective forms of visualization, which allows users to identify, read, or retrieve a value based on different visual encodings. For example, scatterplots colored with additional categorical data were found to be highly effective for comparing individual data and were preferred over dot plots [40].

 **Bar chart.** While bar charts represent information in a way that helps the user identify and read quantitative data easily, the design decisions used play an essential role in user performance. In one study, pictorial bar charts were found to reduce the user performance on retrieving a value task but not beyond the effect already observed based upon their shape [46]. A recent study focusing on a "reading a value" task concluded that there is a need for direct encoding of absolute values and their relationship in bar charts or dot plots [47]. Ultimately, while embellishments are used to improve aesthetics and memorability, they can affect user performance on the task.

 **Parallel Coordinates.** Parallel coordinates are often disparaged as being challenging to use and understand for first-time users. However, one study used the retrieving a value task to show otherwise—the experiment assessed the task using eye-tracking, and it indicated that first-time users quickly learned to use them [48]. The study plotted eight vehicle attributes as axes for 406 cars and used eye-tracking to measure the user's performance on identifying the values in the data.

 **Text.** In visualizations where text is involved, visual complexity affects a user's ability to grasp aspects of the overall structure of the visual display. Optimizing the use of typography in visualizations, e.g., the location of the labeling and parameters including typeface, font size, font weight, color, orientation, intensity (boldness), spacing, case, border, background, underline, and shadow, determines the legibility of text and, thus, influences the understandability of the visualization [49]. One experiment indicated the biases caused by word length, height, and width could impact user accuracy, and the findings provide practical guidelines for improving the user experience [50].

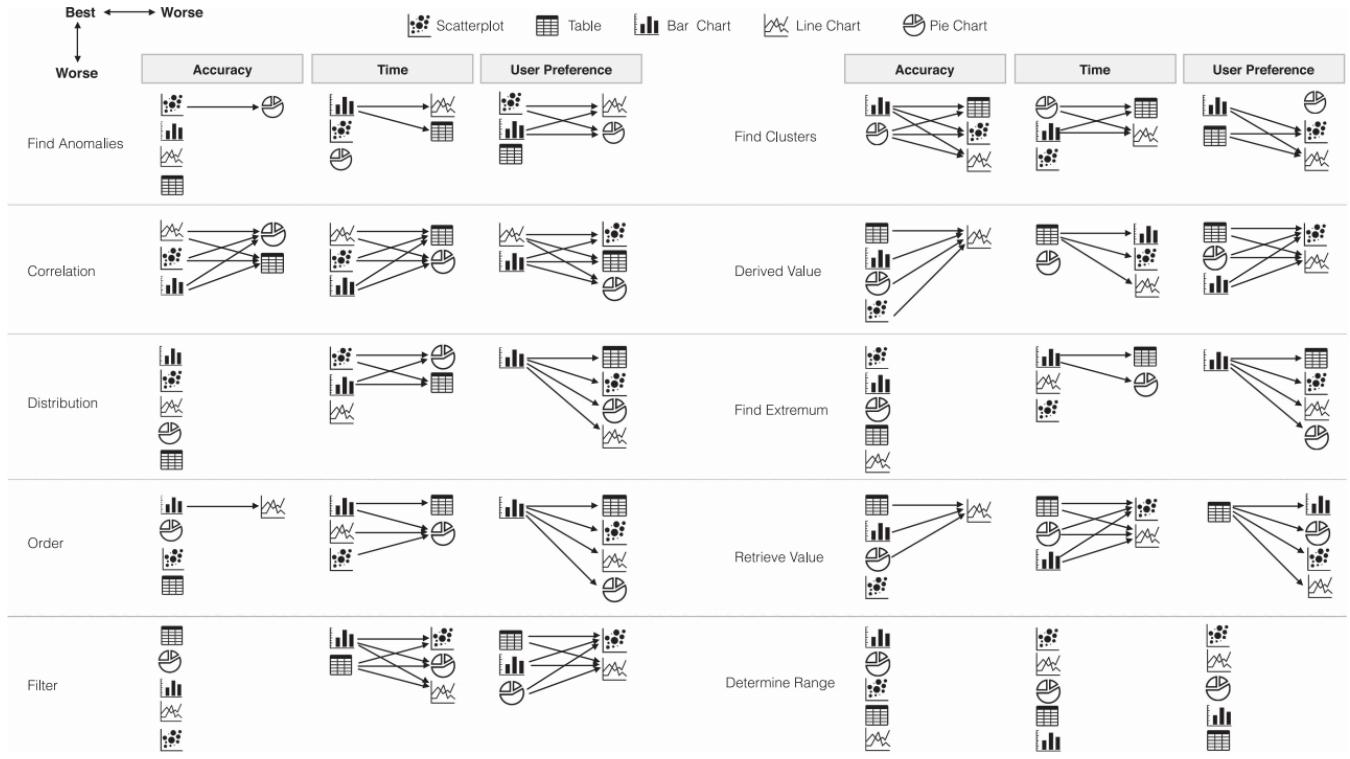


Fig. 5: A recent study evaluated a pairwise relation between visualization types across tasks and performance metrics on small datasets. Arrows show that the source is significantly better than the target. Image reproduced with permission [19].

Therefore, the design of text in a visualization should consider the word typography for any potential biases.

Map. We use larger and higher resolution displays to increase the scalability of visualizations, particularly when data is small, such as in maps. By sheer size, such displays could make it more difficult to attend to the visualization. A perceptual scalability study on maps investigated user performance focusing on retrieving or reading a value task in a larger display [51]. They found that the larger display did *not* result in a time increase or accuracy decrease.

3.1.3 Summary

In summary, we identified several studies using the retrieve value task scattered over various visualization types. As one would expect, user performance varies with the choice of visualization and the visualization design. Prior work indicates that bar charts should be the first choice when accuracy is important. Thoughtful use of embellishments can improve aesthetics and memorability, but they can also affect user performance on the task.

3.2 Filter

 In our search of the literature, filtering essentially came down to two types of tasks. The first, what we would typically call filtering, encompassed eliminating subsets of data used in the visualization. The second was a search task, which focused on finding a target in the visualization.

3.2.1 Visual Encoding

 **Spatial Position, Shape, & Size.** Popout, a pre-attentive effect caused by variations in certain encodings, makes it easy to draw the user's attention to the critical elements of the visualization. A study examined and demonstrated the efficacy of popout in target identification in a group of symbols in a scatterplot using different visual channels, including color, shape, luminance, flashing, motion, and size [6]. On three metrics of performance—perceived success, visibility, and accuracy—shape required high effort with a low performance, whereas motion showed little effort and a high level of performance. The remaining four channels showed a definite increase in perceived visibility and accuracy across intensity.

 A similar experiment investigated searching for a target in a grid matrix using three visual channels—mark size, set size, and color—suggested design guidelines based on grouping, quantity set, and size of visual symbols [31]. While searching for an item was faster when colors were spatially grouped in the grid, the number of symbols had little effect on search time. At the same time, if the number of symbols increased, the performance slowed for random data displays.

 **Color Hue.** Color encoding optimization can increase the effectiveness of categorical data visualization. While categorical colors are easily distinguishable in large color bars or individual plots, when they are inserted into maps or other plots with varying sizes, the variation reduces the visibility of categorical differences. Class visibility is an important measure of how color and spatial distribution of each class affect its perceptual intensity to the human visual system [52].

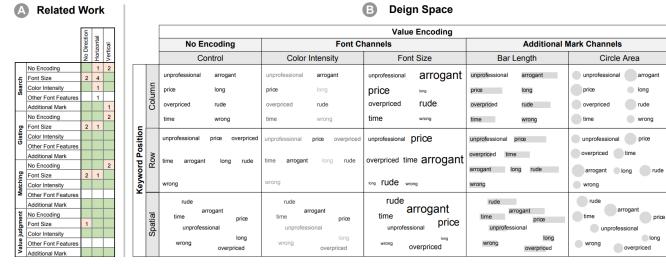


Fig. 6: Different representations affect people's performance in extracting information from text visualizations, such as word clouds. This design space summary shows examples of visualizations generated by different visual encodings and layouts. Image reproduced with permission [57].

3.2.2 Visualizations

Filtering task performance highlights the need to explicitly and directly encode numeric differences between data values [47]. In an experiment with image visualization, finding an image within a set of images showed that both latency and task complexity play a significant role in search behavior [53]. In Saket et al.'s study on small data [19], they found tabular visualizations and bar charts performed well in terms of accuracy, time, and preference on filtering tasks (see Fig. 5).

Parallel Coordinates. In a decade-old experiment, a filtering task was used to assess the usability of parallel coordinates. The performance of the filtering task using eye-tracking showed that users had high fixation time in order to confirm their interpretation of the results [48]. The study showed that the effect of latency in an interactive visual system is more gradual than binary.

Map. Maps are used to represent structures, patterns, and relations in spatial data. Researchers have worked on a variety of map types, e.g., choropleth [54], cartogram [55], geographical map [52], typographic maps [56], etc., to evaluate the human perception in identifying the pattern and structures, and optimize the design process based on the encoding method. As pointed out earlier, class visibility has been studied on maps and shown how color and the spatial distribution can influence perceptual intensity [52]. Filter and search tasks on maps are affected by visual encoding choices, as well as the physical construction of maps. Cartograms were evaluated on the four classes—contiguous, non-contiguous, rectangular, and Dorling—using a qualitative performance analysis [55]. Cartograms that preserve the relative position of the regions in the geographical maps, facilitating faster search, while Dorling and rectangular maps had better accuracy.

Text. The optimal setting of typography parameters determines the legibility of text, thus, influences the understandability of visualizations. In a word cloud study focused on different layout and encodings effects, font size and color performed well compared to encodings using additional marks, such as bars and circles on keywords, in a task of searching for a keyword [57] (see Fig. 6). Furthermore, spatial and column layout of the keywords similarly outperformed the row version of keyword arrangement on keyword searches. A similar study on text confirmed the prior results by showing how searching for target words could be influenced by font sizes [50].

Another study evaluated the effects of four types of layouts, and the results showed that perhaps unsurprisingly, alphabetic ordering was faster for searching for a keyword, order and font size were found to have no impact on searching for a tag belonging to an assigned topic, and words with bigger fonts were more likely to be recalled [58]. This confirmed a prior study that investigated word recall based on classic word clouds with unordered layouts and font size encoding frequency. In a study, participants recalled the words with larger font sizes more often. Their guidelines considered optimizations based on typography (i.e., font weight, size, and color) and word placement (i.e., sorting, clustering, and spatial layout) [59]. In addition to layout and font size, good highlighting mechanisms facilitate search and comparison in textual views or labels. Another study empirically investigated the effective use of highlighting techniques for visualization applications for text data and suggested design guidelines for the effectiveness of nine web-friendly text highlighting techniques [49]. In summary, search tasks on the text and word clouds benefit from the thoughtful selection of size, layout, ordering, and highlight.

3.2.3 Summary

In summary, the filter task is one in which user performance varies based on the design of visualization, choice of visualization, and related set of tasks being performed. Tabular visualizations and bar charts perform well in terms of accuracy, time, and preference. Searching for a keyword in text or a value in a scatterplot can benefit from utilizing encodings font/symbol size and color, as well as alphabetic ordering for text.

3.3 Compute Derived Value

Given a set of data, computing derived values is similar to calculating an aggregate of that data. Many visualization tasks require users to create an aggregated abstraction or statistical summary, which are often referred to as visual aggregations.

3.3.1 Visual Encoding

Spatial Position, Shape, & Size. Cleveland and McGill's seminal work on visual perception utilized the task of computing derived values from the visual representation [3]. In their results, position and length were among the visual encodings with the highest accuracy in quantitative judgment. That study was replicated on a crowdsourced platform using proportional judgment tasks to confirm the effectiveness rankings of visual encoding [33]. A recent study, concentrating on various channels in scatterplots, suggested point size encoding performs well for summary tasks on quantitative encoding [40].

Another experimental study investigated the estimation of average position in line and bar charts [60]. Even though positions are considered a precise form of data encoding, reports of average positions were biased because of underestimation or overestimation of bar positions due to the introduction of a bias called *perceptual pull*—position estimates for a target data were “pulled” toward the irrelevant data in the series.

 **Color Hue & Intensity.** Quite often, colormaps are applied to data, and aggregation tasks are performed. Studies have demonstrated the perception of continuous colormaps is affected by colormap characteristics and spatial frequency. Estimating the values based on colormaps in a continuous quantitative colormap showed no relation between colormaps and spatial frequency, but with increased spatial complexity, estimation error increases [35]. The important point is that spatial frequency impacts the effectiveness of color encoding, but the true impact is task-dependent.

Similarly, another study assessed the efficacy of colormaps for encoding scalar information using binary-choice experiments [61]. Relative distance judgments investigated accessing color similarity between the source and target using reference color. A combination of perceptually uniform color space and color naming more accurately predicted user performance than either alone, but the accuracy was low in both cases.

A study, which used color as an encoding method on time-series data was performed to identify the average in data, and it demonstrated that color-coding techniques showed better average judgments [62]. Symbols are used in a scatterplot to encode values, and lightness is one of the dominant factors of color encoding. A model-based study on computing derived values compared two different types of scatterplots—with a circle as symbols and with spots as a symbol—with varying levels of lightness [63]. The scatterplot with a circle as a symbol performed better in terms of error rate. Gleicher et al.'s large-scale crowdsourced study on identifying aggregate relative judgment to read and the average value in the multi-class scatterplot was another form of computing derived values on various groups of objects [64].

Other. Factors, such as affective priming, social information, or individual user characteristics, directly impact the task performance. Accurate visual judgment is essential to performing the summarization, estimation, or other related computations of a derived value. Harrison et al. performed a series of experiments that showed users articles from the New York Times, followed by a judgment task [21]. The evidence from the experiment showed that affective priming influences the user's visual judgment accuracy. In similar experiments, social information was shown to influence the users on summarization tasks, such as proportion judgment and linear estimation [65]. Other studies have evaluated how individual user characteristics, in particular, working memory (WM) (low, average, or medium), affected the user's performance on visual tasks, such as counting values [66]. The experiment evaluated the task by comparing charts with vertical and horizontal layouts. Users with low visual WM performed faster with a horizontal layout.

3.3.2 Visualization

When comparing the overall efficacy of different visualization methods for computing derived values, Saket et al. showed that for small data sizes: in terms of accuracy table, bar chart, pie chart, and scatterplots were best, in that order; in terms of completion time table and pie chart were best; and finally in terms of user preference, table, pie chart, and bar chart were best [19] (see Fig. 5).



Bar Chart. People prefer linear bar charts because of their familiarity. A comparison-based study between using radial and linear bar charts to visualize daily patterns suggested that 24-hour linear bar charts are more accurate and efficient for summarization tasks [67]. Along with the layout of bars in a bar chart, the design of the bars themselves also affects user performance. In a recent study on comparison tasks on various types of bar charts, two types of compute a derived value task were studied to identify the effects of bar variants on tasks [68]. They found chart design affects the task completion time, and data conditions also influenced the completion time and accuracy.

3D bar graphs are considered to be difficult to understand and generally bad for performing visual analysis tasks. In an experiment where the performance of relative magnitude estimation on pie charts and bar charts with and without 3D was evaluated, the task performance was shown to be better in 2D than 3D [69].

As better design improves performance, visual embellishments are often used to improve aesthetics. Embellishments were evaluated in bar charts on relative magnitude estimations, which confirmed that common embellishments significantly impact the task [70]. In another follow-up study, pictorial charts reduced the user's performance on relative judgments or computing a retrieved value task, but not beyond the effect already observed for their shape [46].



Pie Chart. In an evaluation of pie and donut charts on relative magnitude estimation performance, the baseline donut chart was observed to be as good as the baseline pie chart. Furthermore, user performance on the arc length chart was similar to an area chart, angle pie chart, and angle donut chart [71]. In a similar study, where judgment error in pie chart variations was evaluated, variants of the pie charts, such as a chart with a larger slice, exploded pie, elliptical pie, and square pie caused more significant judgment errors than regular pie charts [72].

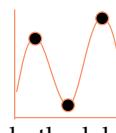


Maps. A study compared different types of cartograms on tasks, such as detecting change and summarizing, found that contiguous cartograms performed better at detecting change, with the lowest error rate and completion time. For the summarize task, contiguous and Dorling cartograms had lower completion times, while the error rate in rectangular cartograms was highest among all four types of map visualization [55].

3.3.3 Summary

In summary, computing derived values is one of the more basic low-level tasks, generally studied as a standalone task. User performance for this task varies greatly based upon visualization and encoding type. Though, the majority of the studies considered only bar charts and line charts. Overall, prior studies have indicated that bar charts, line charts, pie charts, and scatterplots can all be effectively used for summary-based task visualizations.

3.4 Find Extremum



This task focuses on identifying data cases that possess an extreme value of an attribute over its range within the dataset. In this context, the find extremum values task can encompass finding both global and local maxima or minima in data.

3.4.1 Visual Encoding

Position encoded plots, e.g., scatterplots and dot plots, are often used to identify quantitative values [3]. Further, in a study of the effectiveness of visual encodings, the size of marks was more effective than color for quantitative values involving finding or identifying the extremum values [40].

3.4.2 Visualization

For the find extremum task, Saket et al. found all five methods—tabular visualization, scatterplots, bar charts, pie charts, and line charts—performed well in terms of accuracy on small data. Bar charts, line charts, and scatterplots additionally performed well in terms of time. Finally, bar charts performed best in terms of user preference [19] (see Fig. 5). Another recent work compared the performance of design variation in three visualizations—bar charts, line charts, and pie charts—on a task of finding the maximum [73]. The evaluation confirmed the prior hypothesis of better performance in overlaid design versus small multiples and suggested additional new performance improvements.

Bar Chart. Bar charts are often used to read the minimum/maximum quantitative values in data, and recent studies have suggested have indicated that bar design in bar charts influences the user's performance on identifying those values [68]. In another example, a comparison between radial charts and linear bar charts investigated the visualization of 24-hours time. Users had higher accuracy for identifying the maximum value in linear bar charts compared to radial charts [67]. A series of studies were performed for determining the maximum value of delta and mean in bar charts, slope, and donut chart for three variations of their design—small multiples, overlaid, and mirrored [73], [74]. Animation in the design made identifying maximum delta value particularly salient, but the effect did not carry over when the change was large. The identification of maximum mean value held high accuracy for the bar charts with mirrored and stacked bar arrangements.

Line Chart. Line charts are used to visualize time-series data because of the inherent potential of showing the trend. Evaluation of variation in line graph design demonstrates user performance of the finding minimum or maximum values also depends on the design of the visual representation [75]. A modified form of line charts using position encoding, called a “stock chart,”¹ performed significantly better on finding a minimum, whereas the composite line chart² performed well on finding a maximum. However, line charts using a colorfield had the highest accuracy on this task.

Network and Map. A comparison study between node-link graphs and matrix-based representation reported that the task of “finding extremum,” in the form of finding the most connected node, had the highest accuracy and lower completion time for a matrix representation [76]. The results were particularly strong as the size of the graph became large, or the link density increased.

1. The modified stock chart is a line chart with a layering of moving average over the original series to supplement summary judgments [75].

2. A composite line chart is a line chart layered over a bar chart representing averages of discrete subregions [75].

The task of identifying the highest value of an attribute in maps, in the form of top-k identification, a standard cartogram has the highest accuracy when compared to other types of cartograms, i.e., rectangular, non-continuous, or Dorling cartograms [55].

3.4.3 Summary

In summary, finding extremum values is a low-level task mostly studied on bar charts and line charts. Generally speaking, line charts can be used to represent the time-series data when finding an extremum value, while scatterplots and bar charts can be used for quantitative values encoded by position or length.

3.5 Determine Range

 A determine range analysis task has users finding the span of values in a given data for an attribute of interest. The determine range task is another task that has received limited attention.

3.5.1 Visualization

Saket et al.’s evaluation on small data found that for the task of determining a range of values, bar charts had high accuracy, whereas scatterplot performed better on completion time and user preference [19] (see Fig. 5). In an evaluation of multidimensional visualizations, where the task of finding range is evaluated on three user performance metrics of accuracy, completion time, and user satisfaction, parallel coordinates and tabular visualization were found to be effective in terms of accuracy and completion time, but parallel coordinates were least preferred by users [45]. However, this result is only applicable to this limited context.

Bar Chart. A variety of bar charts arrangements designs were investigated on a maximum range task [74]. A stacked arrangement of bar charts gave the highest visual comparison accuracy, while superimposed charts gave the lowest.

Line Chart. For identifying the range of values on time-series data visualizations, line charts based on position encoding had higher accuracy over a modified stock chart, box plot, and a composite line graph, respectively [75]. Though these are variations of the same visualization type, i.e., a line chart, their design variations significantly impact their efficacy.

3.5.2 Summary

In summary, determining a range of values has many unstudied aspects. For example, we did not find any studies which directly investigate the performance on visual encodings. That said, the limited depth and breadth still indicate that scatterplots are preferred when requiring faster performance, bar charts should be chosen when accuracy is needed, and parallel coordinates work well for multidimensional data.

3.6 Sort

 Sorting generally implies ranking the given set of data according to some ordinal attribute. Sorting, as a low-level task, has not received much attention.

3.6.1 Visualization

Saket et al. also studied ordering tasks, which are synonymous with sorting [19] (see Fig. 5). The bar chart stood out in terms of accuracy, timely completion, and user preference. Pie chart, scatterplot, and tabular visualization were next best in terms of accuracy, whereas the line chart and scatterplot were next best for the time completion metric.

 **Bar Chart.** Different types of ranked-list visualizations, such as scrolled bar charts, wrapped bars, piled bars, packed bars, treemaps, and Zvinca plots³, were evaluated in a study that suggested wrapped bars are best for visualizing ranked lists as they provide a simple, compact, and interaction friendly visualization, while treemaps reported the highest accuracy despite the use of area when length is generally a preferable encoding [78].

 **Parallel Coordinates.** An improved form of parallel coordinates, called the progressive parallel coordinates, was studied to understand how data order can mitigate scalability concerns [79]. The application of level-of-detail and randomly accessing individual values were based on an ordering activity.

 **Text.** Sorting a keyword summary alphabetically reduces the time to find a word, and as such, it makes ordered layouts (e.g., column layouts) more effective than unordered layouts (e.g., spatial arrangements) [80]. It was shown that a simple ordering of text data could be much more effective than font size encodings, most notably in searching and retention. In other cases, words can be sorted alphabetically, by frequency, or by a predetermined algorithm. In an experiment on word cloud effectiveness suggesting guidelines for construction, the authors evaluated impression formation and memory by varying font size and layout (e.g., alphabetical sorting, frequency sorting) of words [59]. One important finding was that a list ordered by frequency might provide a more accurate impression as compared to other layouts.

Other. A study focused on user characteristics showed that user performance on sorting values depended upon participants' working memory (WM) (low, average, and medium) [66]. Due to individual differences in perceptual judgments, users with lower verbal WM were slower than others for the sorting task, and users with low visual WM required more time for the task than users with average visual WM. Users with higher cognitive processing speed were also more effective at deriving facts and insights from a visualization than others.

In another study on *perceptual kernels*—a matrix of aggregated pairwise subjective measures of judged similarity—participants were asked to rank the data categories from least to most similar to a target class [30]. This experiment estimated perceptual kernels for visual encoding variables of shape, size, color, and combinations. Based on the judged similarities using Likert ratings among visual variables, findings can be applied to improve visualization design through automatic palette optimization.

3.6.2 Summary

In summary, the sorting, which has many similarities to finding extremum and determining range, has not been studied

3. A Zvinca plot is a layered plot where points replace bars, see [77].

extensively. Nevertheless, studies indicate that bar charts and text-based visualizations are the preferred techniques for quantitative and textual data, respectively.

3.7 Find Anomalies



- The find anomalies task is a form of visual aggregation task, generally involving identifying any outliers or unexpected cases within a given set of data.

3.7.1 Visual Encoding



Spatial Size. With the focus on investigating symbol size discrimination in scatterplots, one study developed a method to pick sizes that are effective for counting outliers and making the judgment as easy as possible [81]. The results show that size perception can be described by the Power law transformation to yield an optimal scale for symbol size discrimination.



Color Hue. Color, motion, and layout (random or grouped) affect the users' attention capacity, and the same has been evaluated on identifying outliers [5]. Search time for an outlier in a matrix form of symbols was shown to vary with color, motion, and layout. Further, color grouped and motion grouped had the best response time, whereas motion random cases were worst in response time. The color group also had the highest identification accuracy.



Color Intensity. Another discriminability study aimed to select luminance levels such that analytical tasks, including counting outliers, were as easy as possible [43]. The performance limit for the task was approached where high-perceived contrast was observed for the data symbols. Additional findings stated that with clustering/groupings of the same data symbols, the task might become easy for low contrast sets.

3.7.2 Visualization

Saket et al. found that for finding anomalies, scatterplots and bar charts performed best in terms of accuracy, time, and user preference [19] (see Fig. 5).



Line chart. Identification of outliers in time-series data using line charts and their variants is complex and reported low user response accuracy [75]. Basic line chart, modified stock chart, composite line chart, and color-field have the same level of accuracy, whereas event stripping outperformed all of these visualizations. The accuracy was low in this visualization for outlier tasks because data spread was confounding as an outlier and vice-versa. Essentially, when data values that would increase or decrease in the opposite direction were focused on for outlier behavior [82].

Other. A recent study on three types of visualization—density plot, histogram, and dot plot—pointed out the outlier detection task as a part of data quality issues [83]. Participants were better able to identify outliers as the flaws in data, but there was no single visualization suggested, which was significantly better among the three.

In an algorithm-based outlier detection, the authors studied outliers and anomalies on box plots and letter-value-box plots [84]. These types of approaches can be paired with visualizations to help analysts explore anomalous data features, especially multidimensional outliers.

3.7.3 Summary

In summary, finding anomalies is defined by the need to identify targets that are different from others in the given set, which varies based upon the visual features, e.g., position, size, orientation, color, and luminance [85]. Scatterplots have been most heavily studied for this task. They are efficient in identifying outliers and detect anomalies in the data. Additionally, line charts easily represent any abnormal or outliers behavior in time-series data.

3.8 Characterize Distribution

This task requires that for a given set of data and a quantitative attribute of interest, the distribution of that data should be characterized over that attribute's value.

3.8.1 Visual Encoding

Characterizing distribution is another visual abstraction task, with a specific focus on pattern or trend recognition. The data can be encoded using several visual features, e.g., position, size, orientation, color, and luminance, which can be crucial in identifying the distribution [85]. For example, Gapminder uses size, position, and color to reveal patterns in global demographics [86], and weather maps use color and orientation to visualize information about wind speed, temperature, and other meteorological data [87].

 **Color Hue.** Colormap design in continuous quantitative maps was used to evaluate the perception of patterns [35]. There was a negative main effect of spatial frequency on pattern perception. Further, the use of low color at low spatial frequency was unlikely to improve pattern perception as compared to a plain grayscale ramp.

3.8.2 Visualization

In Saket et al.'s study, the bar chart was shown to have high accuracy and to be the user preferred method for characterizing distribution tasks, but the scatterplot had better completion time [19] (see Fig. 5). The line chart was next most accurate after the bar chart and scatterplot.

 **Scatterplot.** The scatterplot uses position encoding for the data, and users can identify or detect a data pattern or distribution easily. Varying densities and gaps between data points were shown to influence the user's perception of the distribution or pattern of data points [88]. Further, multiple studies demonstrated the task of characterizing a distribution is influenced by encodings, specifically symbol size and lightness, in a scatterplot [63], [81].

With scagnostics, density is assumed to be a property that shows the concentration of points, which is directly influenced by the distribution of points [89]. This observation led the way to investigating how people interpret trends in a scatterplot and was studied using a sensitivity model. Visual augmentation of scatterplots introduces sensitivity information, which was used to study how people interpret trends in scatterplots [90]. Orientation cues provided by the flow lines give an idea of the local data. Sensitivity or local trends helped in identifying the type of relationship between two variables.



Line chart. Variants of the line chart were evaluated on characterizing the distribution of data, also called the spread [75]. When encodings are position-based, the box plot performed with the highest accuracy, but in the case of color encoding, event stripping performed better. Conventional line charts stood third in the ordering after bar charts and scatterplots for pattern identification. While line charts are usually considered to be the best choice for time series visualization, scatterplots are more effective for showing trends. A study was conducted to merge the two, automatically selecting the right representation for trend exploration in time series data [91]. The choice was affected by the amount of noise, outliers in the data, and aspect ratio. When the noise was small, a line chart is preferred, whereas a scatterplot was preferred with the larger noise.



Parallel Coordinates. A new technique was combined with parallel coordinate, called orientation-enhanced parallel coordinates, to overcome the clutter due to overplotting for characterizing the underlying data distribution [92]. This method improves outlier discernibility by visually enhancing parts of each parallel coordinates polyline through its slope. Interactive evaluation verified the feature of the discernibility of information in complex data.



Maps. Cartograms are popular for geo-referenced data visualizations used to illustrated patterns and trends in the map. Major types of cartograms (e.g., contiguous, non-contiguous, rectangular, Dorling, etc.) were evaluated for comparing trends and analyzed on quantitative performance analysis in terms of completion time and error and subjective preferences [55]. Dorling cartograms performed best on the task involving comparing trends, whereas rectangular performed worst.

Other. Viewers accurately estimate trends in many standard visualizations of bi-variate data. However, visual features, e.g., bias within a bar of visualization, and data features, e.g., outliers, can result in visual estimates that systematically diverge from standard least-squares regression models [93]. Designers should be aware of the distinction between regression by eye and explicit statistical information.

3.8.3 Summary

Scatterplots and line charts have received the majority of attention for the task of characterizing the data distribution. Scatterplots support faster and easier identification of distributions and patterns in data, followed by line charts. A line chart should still be used when the patterns or trends are from time-varying data. Pattern perception on colormaps or chloropleth maps should be sure not to use low color combined with lower spatial frequency. Otherwise, performance will suffer.

3.9 Cluster



Clustering tasks are focused on identifying a similar attribute in a given set of data. A design factor survey study on information visualization defines clustering as a high-level data characterization—"the ability to identify groups of similar items" [94]. Clustering and segmentation of data points in a given dataset reveal characteristics of data and allow visualization designers and practitioners to explore more about the data [95].

3.9.1 Visual Encoding

 **Spatial Position, Shape, & Size.** Symbol or mark size is an influencing visual encoding that affects the density and concentration of point clustering. Additionally, symbol size has a direct influence on identifying the clusters in data, and studies have demonstrated that their discriminability is task-dependent. Li et al. demonstrated the effect of mark size and lightness perception on viewers' ability in multi-class scatterplots for clustering-based tasks [43], [81]. Separability between the symbols or groups of symbols is an important factor in identifying clusters based on the encoding marks. For example, mark shape significantly affects the perception of both size and color, and separability among the three encodings function asymmetrically [96].

Since concentrations of density influence cluster perception, as the size of data points increases, so does the concentration and density. Sadahiro developed a mathematical model to represent cluster perception in point distributions based on proximity, concentration, and density change, and he suggested perception was significantly influenced by the concentration and density change [97]. Varying densities and gaps between groups of points influence the pattern of data points, potentially forming clusters of points [88]. A study focusing on the perceptual optimization of scatterplot design studied standard design parameters, including mark size, opacity, and aspect ratio, and it demonstrated that effective choices of the variables enhanced class separation [98].

 **Color Hue.** Color is another important visual encoding of a scatterplot that influences the visual task of segmentation or clustering and grouping, but how we measure the color difference perception varies inversely with mark diameter [9] (see Fig. 7). Furthermore, the shape of the symbol significantly affects how well we perceive the color difference [96]. Hence, an optimized choice of colors aids users in efficiently understanding separability in multi-class scatterplots. They used a method of color assignment to design scatterplots that optimized class separability perception taking into account density-related factors, such as spatial relationship, density, degree of overlaps between points and cluster, and background color [99], which could not be achieved by the default colormapping.

 **Color Intensity.** Reducing mark opacity can alleviate overplotting to aid various visual analytics tasks, e.g., cluster perception or identification [95], [100], [101] while preserving the spatial information. Different opacity levels aid in enhancing class separation—while low opacity benefits density estimation for extensive data, it also makes locating outliers more difficult [98]. Of course, these solutions have their limits, e.g., when opacity is below available precision, or points are their smallest possible size [102].

Bias. Priming and anchoring effect distorts the user's decision-making process. Deciding the separability of the two clusters depends not only on how far they are apart but also on previously seen stimuli. Valdez et al. elaborated on how repeated exposure to a visualization impacts our interpretation [103]. While the authors presented the biases caused by these effects, at the same time, the results came with the caveat that judgments were not normally distributed, which

further indicating an overestimation of the effect size. The effects can be accidentally caused by chromostereopsis [104] and should be studied on monochrome colors.

3.9.2 Visualization

Saket et al.'s study found that, with small data, bar charts and pie charts outperformed tables, scatterplots, and line charts in clustering tasks [19] (see Fig. 5). The performance in cluster perception in pie charts can be traced back to its effectiveness in facilitating proportional judgments through a part-to-whole relationship [105], [106].

 **Scatterplot.** Scatterplots are widely used to visualize data to reveal patterns of data characteristics such as class segmentation or clusters. Numerous studies on visual cluster separation have been conducted to identify the effects of various design factors and visual encodings.

A qualitative evaluation of cluster separation measures study suggested taxonomy of four factors—scale, point distance, shape, and position—that influence separation perception [36]. With the focus class separation, Sedlmair and Aupetit's extended their prior work and evaluated 15 state-of-the-art class separation measures in a study aimed at mitigating human judgment impacts. They rely on human ground truth as input to a machine learning framework that was used for evaluating the quality of dimension reduction [107]. A further continuation of the work included even more measures for improved matching to human perception [108]. The ScatterNet method captured perceptual similarities between scatterplots by applying a deep learning model that was designed to emulate human clustering decisions [109]. The scagnostics technique focused on identifying the patterns in scatterplots, including clusters [110], [111]. However, a recent study showed that scagnostics do not reliably reproduce human judgments [112]. The commonality in all of these studies is that they are algorithmically oriented, and most of their evaluations did not consider visual channels.

Sadahiro presented a mathematical model that suggested perception is significantly influenced by the concentration and density change [97]. ClustMe used visual quality measures to model human judgments to rank scatterplots [113]. ClustMe performed well in reproducing rational decisions for cluster patterns. Quadri and Rosen built and tested a topology-based model of human perception of scatterplots that considered data distribution, number of data points, size of data points, and opacity of data points in cluster perception [101]. Their model demonstrated the strong relationship between all of these factors and the perception of cluster separation.

 **Parallel Coordinates.** Parallel coordinates are useful for a variety of tasks, including clustering in real-world applications [114]. One study that measured the participants' learning outcomes used clustering as a primary task [115]. The results showed a more engaging experience for interactive parallel coordinates than static, and the users did not find it challenging to learn the data item mapping to the parallel axes.

Orientation-enhanced parallel coordinates were developed especially for large datasets to improve the display of emphasizing the underlying data structure, such as allowing the discernibility of clusters [92]. The evaluation demonstrated that orientation made it easier to identify the

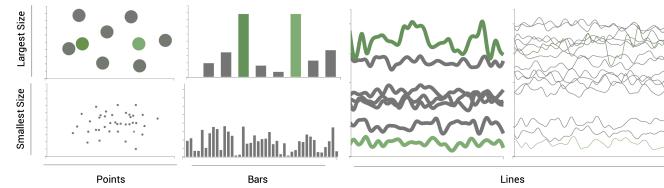


Fig. 7: The perceived color difference varies inversely with size, and elongated marks provide significantly higher discriminability. Colors on longer marks were also more discernible than shorter bars of equal thickness. The perceptible color differences for lines vary inversely with thickness. Finally, perceptible color differences for points vary inversely with point diameter. Image reproduced with permission [9].

data clusters between the two data dimensions and allowed multiple small clusters between the first two dimensions to be visually enhanced. Ultimately, there remains some question as to whether scatterplots or parallel coordinates are better for identifying clusters, as one study found that parallel coordinates better showed the actual shape of clusters [116].

Networks. Graph layout algorithms optimize visual characteristics of visual encodings to create intuitive visualizations. The layout of graphs was investigated in an evaluation study where participants produce their graph from a graph shown earlier to depict clusters [117]. Users reach confidence and higher overall task accuracy in visualization during the interaction when rapidly adjusting the visual encodings.

Text. A study to investigate word recall memory, based on classic word clouds layout and font size, stated that word placement (i.e., sorting, clustering, spatial arrangement) is an important consideration as it affects the users' recall potential [59].

3.9.3 Summary

As scatterplots demonstrate the clustering of bivariate data effectively, most studies we identified used scatterplots. In the case of multivariate data, parallel coordinates and scatterplot matrices can be used to visualize data.

3.10 Correlate

 Correlation, in general, is a relationship between values of two or more attributes in a dataset. Formally, correlation is a statistical measure of the linear relationship between two quantitative variables, represented by a correlation coefficient. A positive correlation indicates the extent to which those variables increase in parallel, and a negative correlation indicates the extent to which one variable increases as the other decreases.

3.10.1 Visual Encoding

Evaluating the correlation perception helps identify people's abilities to perceive and judge differences in visualizations. Numerous studies have focused on identifying linear correlation in a data distribution with visual encodings, including the slope of the points, marker size, opacity, and color [27], [98], [118], [119], [120]. Rensink and Baldridge showed that perception of correlation in scatterplots could be mathematically modeled using the perceptual laws of Weber's law [118].

Further, Chang et al. [121] also use Weber's law to provide a guide for practitioners to select a visualization.

3.10.2 Visualization

Identifying correlation supports the decision-making process in multivariate data visualization. A comparative study of parallel coordinates, tabular visualization, and scatterplot matrix reported tabular visualization and the scatterplot matrix have high accuracy, whereas parallel coordinates stand out in terms of completion time [45].

The Harrison et al. study compared and ranked scatterplots to other visualization methods (parallel coordinates, stacked area charts, stacked bar charts, stacked line charts, line charts, ordered line charts, radar charts, and donut charts) for correlation perception tasks and stated that Just-Noticeable Difference (JND) in correlation could be modeled by Weber's law [27] (see Fig. 8). Kay and Heer proposed a log-linear-based re-analysis of the results of Harrison et al.'s work [122] (see Fig. 9). The critical finding of these works was that scatterplots stand above all other tested visualizations for precision on both positive and negative correlations.

As a counterpoint, Saket et al.'s reported that line graphs stand out for detecting correlation in terms of accuracy, time, and user preference, followed by scatterplots [19]. However, the experiments only used a small number of points, adding further complexity to the discussion of which method is truly most effective for identifying correlations. Nevertheless, the results are supported by earlier research reporting the effectiveness of line charts for trend-finding tasks [123].

 **Scatterplot.** Correlation is one of the more extensively studied tasks in this taxonomy, particularly concerning scatterplots. One of the first perceptual studies on correlation on scatterplots by Bobkoff and Karren [124] formed the basis for many other experiments. They measured a direct estimation of Pearson's product-moment correlation coefficient, like many other correlation perception studies [119], [125], [126], [127], [128]. However, Sher et al. conducted a study about measuring the offsets of human perception of correlation when changing visual variables, and they found that humans perceive correlations differently compared to the statistical measure of Pearson's product-moment correlation coefficient [129].

	$r = 0.1^*$	$r = 0.3$	$r = 0.5$	$r = 0.7$	$r = 0.9^*$	overall
pop-negative	pop-negative	scatterplot-positive	scatterplot-negative	scatterplot-negative	scatterplot-positive	scatterplot-positive
scatterplot-positive	scatterplot-positive	pop-negative	scatterplot-positive	scatterplot-positive	pop-negative	pop-negative
scatterplot-negative	scatterplot-negative	scatterplot-negative	pop-negative	pop-negative	scatterplot-negative	scatterplot-negative
stackedbar-negative	stackedbar-negative	stackedbar-negative	stackedbar-negative	ordered line-positive	stackedbar-negative	stackedbar-negative
ordered line-positive	donut-negative	ordered line-positive				
donut-negative	donut-negative	donut-negative	donut-negative	donut-negative	ordered line-negative	donut-negative
stackedarea-negative	stackedarea-negative	stackedarea-negative	stackedarea-negative	ordered line-negative	stackedbar-negative	stackedarea-negative
ordered line-negative	ordered line-negative	ordered line-negative	stackedarea-negative	stackedarea-negative	stackedline-negative	ordered line-negative
stackedline-negative	stackedline-negative	stackedline-negative	stackedline-negative	stackedarea-negative	stackedarea-negative	stackedline-negative
pop-positive	pop-positive	pop-positive	pop-positive	radar-positive	pop-positive	pop-positive
radar-positive	radar-positive	radar-positive	radar-positive	radar-positive	pop-positive	radar-positive
line-positive						

Fig. 8: Harrison et al. leveraged perceptual laws to evaluate and rank the effectiveness of visualizations for representing correlation. One interesting finding is that judgment precision had a striking variation between negatively and positively correlated data on certain visualizations, e.g., parallel coordinates. Image reproduced with permission [27].

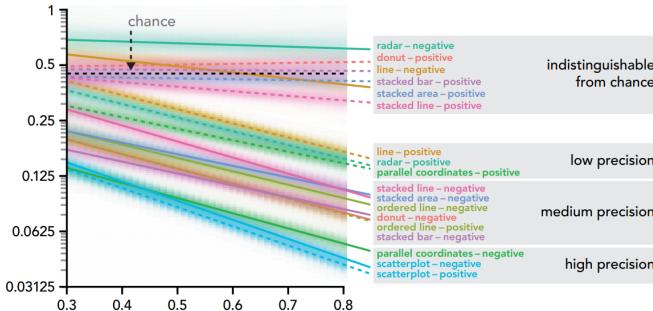


Fig. 9: The study by Kay and Heer presented a series of refinements to the model presented by Harrison et al. [27] (see Fig. 8), including the incorporation of individual differences, log transformation, and Bayesian modeling. The left side shows the posterior probability distribution over the mean log (JND) for each value of r using the Bayesian censored log-linear model. The right side shows the ranking and grouping of visualizations based on how precise people's estimations of correlations are (lower JND implies higher precision). The new model demonstrated notable differences, e.g., parallel coordinates–negative and scatterplot–negative swap positions. Image reproduced with permission [122].

Correlation was also among the first visualization tasks to be studied using JNDs in four experiments [4]. The study found that users become more confused with low correlation values than with high correlation values. The reason behind the effect was that the user's judgment of discriminability increased with the increased strength in the variable association. Rensink's perception of correlation study opened a new door to effective visualization design by demonstrating a precise way perception of correlation in a scatterplot could be modeled using Weber's law and JND [118]. Their design studies were focused on the comparison of performance using dot size and density.

Best et al. investigated correlation coefficient estimation in the laboratory-based study where they found that human brain activity during correlation perception increases as correlation is decreased [130]. Findings indicated that different relationships on a scatterplot are processed differently. However, perceptually, scatterplot processing was similar, and participants used visual features to code the pattern.

A recent comparative survey included all correlation studies in a timeline taxonomy and investigated a hypothesis—viewers attend to a small number of visual features, e.g., shape, dispersion, and orientation of scatterplots to discriminate correlation in scatterplots [22]. The report compares the findings with previous pivotal correlation perception studies using Weber's law. Scaling is another shape property that plays an important role in identifying the correlation between the variables. An early study by Cleveland et al. reported that variables in a scatterplot look more correlated when scales are increased [119].

Map. Visualizations on maps show spatial data, and design variations are useful in identifying correlation over geo-temporal data variables. Map lineups used JNDs for evaluating correlation on choropleth maps [54]. They model the differences in visual stimuli of map color and are useful in controlling spatial auto-correlation and

increase user's confidence. The comparison of average JNDs measured across all three geographies demonstrated that JND increases and becomes more noticeable when irregularities become more regular.

The encoding choice on the geophysical maps has a direct influence on correlation judgments. A study on geotemporal visualization investigated the task of identifying the correlation between two variables that evolve over time and space [131]. The findings showed that the design choices of geo-temporal multivariate data visualization would impact how users detect a correlation between variables over space and time. The vital design guidelines from this study are: (1) small multiple visualizations on maps are better for identifying correlation at a specific point in time; and (2) for identifying correlation for time steps on single maps, bar charts are better than other choices.

Other. Design variation, such as small multiples, overlay, and mirroring on three visualizations (bar, slope, and donut charts), were shown to influence the correlation task [73]. In the study, participants struggled to use motion animation to extract and compare the correlation between data sets, whereas mirrored arrangements over adjacent arrangements achieved precise results. The comparison arrangement design evaluation confirms the prior hypothesis of better performance in overlaid designs versus small multiples.

3.10.3 Summary

The majority of the correlation studies have been conducted on scatterplots, but recent works have also diverged towards other visualization types. Studies have by-and-large shown that either scatterplots or parallel coordinates should be used for correlation tasks. However, their performance may interchange with positive or negative correlation due to the different representations of positive and negative correlations in parallel coordinates. For spatial data correlation choropleth maps can be an effective approach.

3.11 Compare

Compare is a compound task that was mentioned as an intentional "omission from the taxonomy" of Amar et al. [11]. The task of comparison often involves another subtask, e.g., retrieve a value, compute a derived value, etc., followed by comparison operation. Comparison was implicit in many of the prior tasks. For example, the find extremum task (Sect. 3.4) often requires comparing a set of candidate values to the rest of the data, e.g., "which cars are more fuel-efficient, Japanese cars or American cars?" [11]. This section focuses on the performance of comparison tasks based on different visual features in their design of the same visualizations to analyze the perception judgment. Based on the quantity and importance of studies of this type, we have included it as the eleventh task of the taxonomy.

3.11.1 Visualization

Bar Chart. The majority of the comparative study performed on bar charts investigate the user performance with design variations. The visual comparison task can be traced back to Cleveland and McGill's work, where charts are shown, and different bar chart

designs impact the accuracy and comparison of the viewer's perceptual task [3]. The study reported that a comparison between adjacent bars is more accurate than between widely separated bars.

The findings from that study were extended to four different types of bar charts [20]. Some of the critical findings on bar chart comparisons are: short bars are difficult to compare; the gap between stacked bars can prevent part-to-whole comparison errors; distractors in bar effects unaligned bar comparison; and separating bars in space makes the comparison of their height more difficult. At the same time, comparing the variants of the bar chart design provides different levels of performance on completion time and standard error [68]. Variants of bar charts were compared based on the visual features in their designs, and their perception judgments were used to discover error and distraction factors. Charts with different overlays or hybrid designs that combine aspects of juxtaposition and explicit encoding with superposition are just as good or better than sole juxtaposition or explicit encoding-based charts on individual tasks. A comparison helps the designer to identify differences in the representation of data.

Other forms of design variation in bar charts are embellishments which attach aesthetics to visual display. Different bar charts with visual embellishments were compared to evaluate the accuracy, and the results demonstrate that accuracy was not worse, but at the same time, it did not provide better outcomes [70].

Comparing a visualization to a mental image is similar to statistical analysis, and thus repeated interpretation of visualization is sensitive to the multiple comparison problem—the probability of discovering false insights when visualization is examined more times or compared [132]. The study was performed on different types of bar charts to measure and compare the accuracy of user-reported insights such as shape, mean, variance, correlation, and ranking.

Another work on bar charts is compared with Microsoft Excel to show the difference in activities in terms of sequences of action and pipelines [133]. Participants tend to follow a linear pattern when using Excel, whereas while using tiles, they followed a cyclical pattern. In a study of five types of ranked list visualization, a comparison of two data items takes the longest time with the lowest accuracy on Zvinca plots [78]. A treemap outperformed all other visualizations on accuracy.



Line Chart. A series of studies have evaluated and compared the performance of line chart variants: horizon graphs⁴ [32], [134], [135], [136], colorfields, [62], [137] braided graph [134], and small multiples [138]. A recent study assessed horizon graphs and colorfields, along with a simple line graph on the perception of similarity between time series—two patterns were considered similar irrespective of their amplitudes or their stretching along the time dimension [137]. Layered bands are more useful as chart size decreases.

In another study, where mirrored and offset horizon graphs were compared, found that estimation error and time increased at four bands in the horizon graphs [32].

⁴ Horizon graphs split the space (mainly vertically) and attempted to optimize the vertical footprint to visualize multiple time series [134].

At the same time, the different chart types did not affect the estimation time or accuracy. The effects of chart size of horizon graphs and layering on comparison and estimation showed that horizon graphs performed better than line charts for small chart sizes. A time-series visualization comparison studied line charts, horizon graphs, and color fields in a similarity perception comparison where the choice of visualization affects the viewer's perception of temporal patterns [137].



Network. Complementary views are best and increased the accuracy of network exploration tasks.

Chang et al.'s work compared matrix diagrams, node-link diagrams, and weighted networks to find effective matrix representations in side-by-side views for network exploration tasks on error, completion time, and user preference [139]. Findings state node-link and matrix views are well suited for different visual tasks. Another study compared graphs either for different tasks or different datasets to measure their effectiveness. Graph edge attributes with uncertainty were visualized using two separate visual variables. For the task of comparing two graphs on overall strength or certainty showed that lightness was an effective mechanism for uncertainty [140].

Readability is another network feature evaluated in a comparison study between node-link diagrams and their matrix-based representations on generic low-level tasks [76]. They found readability depended upon graph detail, familiarity, graph meaning, and the layout used to visualize them, but the findings also reported that it deteriorated when the size of graphs and link density increased. In a recent work on the perception of graph properties, three layout algorithms—force-directed, multidimensional scaling, and circular—were modeled and compared using Weber's law to discriminate between graphs [141]. The results showed that Weber's law could be used to model density perception.

3.11.2 Summary

For the comparison task, we observed a relatively fair distribution of studies on bar charts, line charts, and scatterplots. Ultimately, the choice of visualization is more subtask-driven and design-dependent.

4 DISCUSSION

We have presented a systematic review of research in the perception of visualizations. One important observation from our survey is that much of the research has been pursued through the lens of low-level tasks in terms of efficiency and effectiveness. Early works were broad in nature. For example, Cleveland and McGill's early work demonstrated how the many different types of visual encodings influence perceptual judgments [3].

4.1 Visualization Effectiveness Is Task-Dependent

One of the important conclusions we have seen time and again in prior work is that low-level task effectiveness varies with the dataset at hand, the visualization used, and specific design variations with the visualization. While some visualization types tended to perform better than others on average, it seems, from our observation, there is no single

visualization that is suitable for all situations. Saket et al. seemed to agree in their study of five visualizations on small datasets using ten low-level tasks [19], when they stated that “No One Size Fits All.” In other words, depending on the task at hand, various visualizations perform can perform better or worse. In a similar vein, even within a single visualization type, design variations can have a serious effect on performance. Mylavarapu et al.’s study on ranked-list visualizations, which included wrapped bars, packed bars, piled bars, and Zvinca plots, quantified the differences and trade-offs for three tasks. The effectiveness of the representations varied, as each had its own strengths and weaknesses that depended upon the task, data, and user [78].

4.2 Progress Understanding Graphical Perception

Investigating and evaluating the effectiveness of visualizations to optimize their visual design is a perennial topic. We witnessed the continuously evolving nature of perceptual research, with the majority focusing on visual task judgment. The recent upward trend in perception-based visualization studies, as noted in Fig. 1 and supported by two recent related STAR reports by Borgo et al. [13] and Behrisch et al. [17], show the maturing of this area of research. Many of the works discussed in this report have applied perceptual laws to evaluate visualizations, hopefully leading to better visual design.

Much of our knowledge about perception in visualization is taken for granted, and despite the diverse set of perceptual research in visualization, many topics we have presumed as “fact” have never been sufficiently studied. Take, e.g., parallel coordinates—despite their reputation, studies have shown that parallel coordinates may not be as difficult to use as we think [48]. The point here is not entirely to discard the facts but instead to consider that re-evaluating what we know might lead to better design guidelines. This align with some of the work of Kosara [142], where the effort was to tease apart what we know and what we only think we know, using examples. The goal is to point out specific gaps in our knowledge and to assist researchers in starting investigations of the underlying theories to systematically build up a better foundation for our field.

As the research objectives and methodologies have evolved, insights have become more fine-grained and nuanced, e.g., Chung et al. observed that color saturation with size could be used as an ordering variable [37]. Similarly, Szafir’s work measured perceptual judgments on color difference that varied with size encodings, e.g., bar-width, circle-radius, line-width [9]. These innovations, while enlightening, also stand somewhat in contrast to applied works, e.g., Colin Ware’s book [2] on visualization design and perception, which provides a practical view of effective design.

4.3 Limitations of Scope and Reproducibility

As findings become more nuanced, so too does the scope of the findings. Most individual studies come with limitations in scope, complexity, objective-goal, and sometimes with datasets or demographics. For example, studies that are performed using limited and less diversified subjects, e.g., on computer science students, need to be further investigated before their guidelines could be put into practical use for a

broader population. Readers and practitioners need to pay close attention to the limitation studies, as they are critical in effectively applying their conclusions.

An integral part of understanding the limited scope of studies is considering the issue of reproducibility in studies. Reproducibility is being increasingly encouraged by communities of science to validate conclusions and to extend, re-evaluate, or specialize the original ideas [12], [143]. With a few exceptions, e.g., [33], [122], there has not been significant consideration of replication within the perception studies we have surveyed. Several fields of science are experiencing a “replication crisis” that has negatively impacted their credibility. The problem is further exacerbated by the limited scope of visualization studies, specificity of experiments, and difficult logistics required for reproducing a study.

The lack of sufficient incentives, e.g., generating highly cited publications at high-quality venues, for reproducing studies also stands in the way. Recent suggestions for mitigating this problem revolve around tightly coupling original research with replication studies [12], [143], though this approach does not address the lack of incentives. One recent step in the right direction has been the trend of encouraging authors to make their research data available. A great set of recommendations are provided for open research practices in [144].

4.4 Limitations and Open Questions

Throughout this survey, we uncovered various open challenges and areas requiring more study.

Beyond Scatterplots, Bar Charts, and Line Charts. Many research studies in visualization are driven by popularity, familiarity, and applications. Table 2 provides an overview to what extent the various areas have been studied in our survey of papers. Clearly, scatterplots, bar charts, and line charts received the majority of attention in research papers. It is undeniable that bar charts and line charts are widely used to represent univariate data, while scatterplots are used to represent bivariate relationships effectively. Parallel coordinates are an interesting case—even though there has been extensive design-based research work on parallel coordinates, the number of perceptual evaluations is low. One reason may be that it is complex and challenging to perform the perceptual evaluation. Similarly, we observe a bias in Table 2 to the tasks of Retrieve Value, Compute Derived Value, and Compare. While it is unlikely that the distribution of studies will balance out, insights gained from some of the less frequently studied techniques can have significant value. Researchers should consider focusing their new research efforts on the less studied tasks and visualization types.

Limitation of Experiment Data Collection. One topic we did not address in any detail is the collection of human subject data used in these studies. Variations in data collection make meta analyses difficult and can harm reproducibility by fixating on limited subject pools or specific experimental setups. The evaluation and perceptual studies we identified most often used metrics such as accuracy, time to completion, and subjective preference. Data collection methods range from keyboard input to mouse input to eye-tracking (e.g., scanpath and fixation) to voice. The experimental strategies also varied from methods of adjustment or interaction, forced-choice, and think-aloud protocols, which can provide more

TABLE 2: Table summarizing the number of studies we reviewed per task, visual encoding, and visualization type. The table reveals which area have received the most attention and those that need more work.

Task	Visual Encoding				Visualizations										Total	
	Spatial Position & Shape	Spatial Size	Color Hue	Color Intensity	Scatter-plot	Bar Chart	Line Chart	Map	Network	Heat-map	Bubble	PCP	Pie Chart	Text	Other	
Retrieve Value	22	21	14	14	7	9	5	7	1	1	4	3	3	3	11	31
Filter	7	9	6	4	3	3	1	1	0	2	0	4	1	4	6	15
Compute Derived Value	38	36	14	20	4	20	10	5	2	2	4	2	4	2	16	48
Find Extremum	15	14	9	9	2	7	6	3	1	3	0	2	1	1	7	19
Sort	9	10	5	6	5	6	4	1	0	0	1	2	1	1	7	12
Determine Range	11	10	6	5	5	2	2	2	0	0	1	2	1	1	7	14
Characterize Distribution	14	11	12	13	7	2	5	0	1	2	1	2	1	1	6	19
Find Anomalies	5	8	4	5	5	1	3	2	0	1	1	0	1	1	3	10
Cluster	9	8	5	5	7	1	1	3	1	2	2	1	1	2	3	13
Correlate	14	10	4	2	11	2	3	1	0	0	0	5	1	1	6	17
Compare	24	22	10	18	6	10	7	4	4	5	1	2	1	2	9	34
Total	94	84	48	57	42	38	27	22	13	12	8	8	7	7	40	

cognitive and problem-solving insights. Many of the experiments also suffer from low ecological validity, which refers to how closely the experimental setting matches the real-world setting in which the results might be applied [145]. There is an explicit trade-off between experimental control and ecological validity [146]. Studies with high ecological validity closely reflect real-world use, while those with low ecological validity are often highly controlled. Finally, we found variation in the participant pools used, generally consisting of university students or random subjects on crowdsourcing platforms. All of these variations in metrics, apparatus, and subject pools limit the applicability studies. For example, familiarity with the visualization, task, datasets, and design can be extremely biasing to participants' performance. In the same regard, other biases, such as affective priming, can influence perceptual judgment. These limitations require serious consideration when applying conclusions to designs.

Relationship to Cognitive Processes. Perception and cognition are tightly linked and difficult to separate. Almost every study we discussed had some cognitive component to it. For example, estimating correlation is both perceptual and a cognitive task. In terms of the efficacy of visualization, cognition plays as important of a role as perception does. For example, one of the metrics on task performance, which was largely ignored or at least largely unreported in experiments, was confidence. Based on the complexity of the task, e.g., interaction versus forced-choice, visualization type familiarity, e.g., parallel coordinates, comfort with the type or size of data, or the type of experimental study being performed, subject confidence can vary widely. There is even a possibility that the user's confidence will vary task-by-task. A related challenge is finding the right set of subjects, which is a particular limitation when recruiting participants through crowdsourcing. Due to the population diversity, their experience and confidence vary, and it is difficult to check quality. It fits then for researchers to evaluate the participants' level of confidence in task judgment, and accordingly, data quality should be evaluated. New studies should consider the participant's level of confidence and how this affects high- and low-level tasks.

4.5 Conclusion

We have presented our report with a particular focus on the links between visualization types, visual encodings, and tasks. Through our taxonomy, we wanted to emphasize

perception-based research findings and their impact on visual encoding and visualization choices. We believe this report will be a valuable starting point for those designing visualizations and researchers looking to advance the state-of-the-art in perception-based visualization research. It should be noted, however, that the aim of this survey was not to directly summarize the visualization recommendations but instead to provide a broad understanding of the topics that have been evaluated. Furthermore, all of the perceptual studies were studied under limited conditions and come with caveats that we could not enumerate.

5 APPENDIX: PERCEPTION FUNDAMENTALS

We provide a brief introduction to some fundamental concepts of perception discussed in the paper. The presented information is concise, and we encourage readers to refer to the related articles for more in depth information.

5.1 Psychophysical Effects

Psychophysics is a set of methods relating sensations to the characteristics of a stimulus. It is used to quantitatively investigate relationships between physical stimuli and the sensation of the perception they produce [147], [148]. The transition of psychology from a philosophical to a scientific discipline was facilitated when G.T. Fechner introduced techniques to measure mental events. The attempt to measure sensations through the use of Fechner's procedures was termed psychophysics and primarily investigated the relationships between sensations in the psychological domain and stimuli in the physical domain. Central to psychophysics is the concept of a sensory threshold, that measurement can have a differential and an absolute sensitivity. The *absolute threshold* or *stimulus threshold* is defined as the smallest amount of stimulus energy necessary to produce a sensation. The *differential threshold* was defined as the amount of change in a stimulus required to produce a just noticeable difference (JND) in the sensation.

Weber's law, or Weber-Fechner's law, relates to psychophysics and is used to determine the relationship between the perceived change in a stimulus and the actual change, which has been used to modeling how humans perceive certain features in a visualization. The law states that the change in a stimulus that will be just noticeable is a constant ratio of the original stimulus [149], [150].

As an example of Weber's law, consider the act of lifting a 5 kg weight. Adding a small amount of weight, say 0.1 kg, may not make the weight *feel* any heavier. With further additions of weight, the difference will eventually be noticeable. Weber's law is the ratio of change in the stimulus ($\Delta I = 0.1$ kg) to the stimulus magnitude ($I = 5$ kg), which is 0.02. Weber's law has been shown to hold for weight discrimination, visual discrimination, and tone discrimination [151].

Just-Noticeable Difference (JND), also known as the difference threshold, is the minimum level of stimulation that a person can detect, usually 50% of the time, though other ratios can be used [152]. For example, one is asked to hold two objects of different weights—the JND would be the minimum weight difference between the two that one could sense half of the time. JND is used as a component for the perceptual studies, which are required to determine how much a given stimulus must be regulated in order for a human to detect a change reliably. The relation between JND and the stimulus can be represented using Weber's law [149], [150] as follows: $dP = k \frac{dI}{I}$, where, dP is the differential change in perception; k is the Weber fraction; dI is the differential change in stimulus, and I is the actual intensity of the stimulus. With the given I and Weber fraction, JND corresponds to the minimum change of stimulus will produce a noticeable difference in the perception. The application of JNDs with psychophysics evaluation is useful for measuring human judgments in the effectiveness of visual encodings for different tasks or design improvements.

5.2 Bias and Effect

In the process of evaluating visualizations, the effect of bias, whether cognitive or perceptual, is critical to understanding and evaluating experimental results. The primary types of bias observed and studied in visualization are perceptual biases and cognitive biases.

5.2.1 Perceptual Bias

Perceptual biases are systematic errors that occur at the perceptual level in perceiving visualization and/or related tasks. Some types of perceptual biases studied include clustering illusions and priming biases [153]. For example, the clustering illusion is where people underestimate the variance seen in patterns in a small set of random data [154], [155]. Priming relates to associative memory theories where a concept/effect is quickly activated after a similar concept/effect has been activated.

Perceptual bias has been studied in several regards. Change blindness refers to the inability of humans to recognize large visual changes between images. An optimization-based method introduced and evaluated an approach to generate "spot-the-difference" alternatives [156]. Perceptual biases have been studied in virtual reality platforms, e.g., in the perception of size and stiffness of virtual objects [157] and how photorealism negatively affects our perception of virtual characters [158]. Irregularities in data can cause bias, influencing a user's response to the conclusion of analysis [83], [159]. One study found that perceptual biases influence a user's awareness of uncertainties, further influencing the user's trust building [160]. Finally, researchers have investigated

how bias and perception intersect to create deceptive views. Pandey et al. developed a method to quantify and compare the exaggeration caused by misleading representations [161].

5.2.2 Cognitive Bias

Cognitive bias research has grown considerably at both the cognitive science level and specifically for visual analytic and decision-making tools. Cognitive biases differ from perceptual biases in that they persist even if the information has been correctly processed at a perceptual level. Broadly, collective works on cognitive biases can be found in [162]. The following are examples of cognitive bias topics that have been evaluated in visualization.

Attraction effect is a type of cognitive bias where the presence of irrelevant alternatives influences the choice between two options. When the user's choice between two options is influenced by the presence of an irrelevant (dominated) third alternative. A good example, coming from [164], involves selecting between two candidates based on a rating of their education and crime control plans (see Fig. 10). Bob has a solid education plan, while Alice's strategy for crime control is excellent. The choice is difficult if we consider both the criteria. A third candidate, Eve, called the "decoy," is focused more on crime control than education, though not as good as Alice. The introduction of Eve as alternative biases our preference towards Alice. Awareness of the attraction effect is important, as it may introduce skew into decision-making tasks, as evidenced by Dimara et al.'s study on scatterplots [163].

Anchoring effect is a type of cognitive bias, where a stimulus might influence human judgment at the perceptual level of the decision-making process. Anchoring effects and ordering effects describe how the order in which information is presented can affect the perceived size of an effect, with subjects across a wide range of domains tending to assign more rhetorical weight to evidence that comes near the beginning of a sequence.

Priming is another form of stimulus that can influence human judgment. The priming effect is a phenomenon in which an alternative perceptual stimulus influences human responses [165]. Priming effects are seen more frequently than anchoring effects in separability judgment.

Emotions, in psychology, are defined by two dimensions: *valence*, positive or negative feelings, and *arousal*, the intensity of the feelings. **Affective priming** is the technique of inducing emotion, also known as *affect*, in a user to study the impact on cognitive tasks. It involves manipulating valence and/or arousal via emotional stimuli. Harrison et al.'s contribution to affective priming suggested that it can influence accuracy in graphical perception tasks [21], [166]. The crowdsourced study indicated that affective priming significantly influenced visual judgment, while positive priming improved accuracy.

	Bob	Alice	(Eve)
education	★★★★★	★★	★★
crime control	★★	★★★★★	★★★★

Fig. 10: Example of attraction effect in selecting hypothetical election candidates. Image reproduced with permission [163].

The cognitive biases of both anchoring and priming suggest that the decision-making process not only depends on what the visual features currently look like but also on the previous frame of reference.

ACKNOWLEDGMENTS

We thank our reviewers for their helpful feedback on the construction of this paper. The project is supported in part by the National Science Foundation (IIS-1845204).

REFERENCES

- [1] M. Tory and T. Moller, "Human factors in visualization research," *IEEE Trans. on Visualization and Computer Graphics*, vol. 10, no. 1, pp. 72–84, 2004.
- [2] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.
- [3] W. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, 1984.
- [4] M. Doherty, R. Anderson, A. Angott, and D. Klopfer, "The perception of scatterplots," *Perception and Psychophysics*, 2007.
- [5] S. Haroz and D. Whitney, "How capacity limits of attention influence information visualization effectiveness," *IEEE Trans. on Visualization and Computer Graphics*, vol. 18, no. 12, 2012.
- [6] C. Gutwin, A. Cockburn, and A. Coveney, "Peripheral popout: The influence of visual angle and stimulus intensity on popout effects," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2017, pp. 208–219.
- [7] C. Healey and J. Enns, "Attention and visual memory in visualization and computer graphics," *IEEE Trans. on Visualization and Computer Graphics*, vol. 18, no. 7, pp. 1170–1188, 2012.
- [8] R. Rensink, "Attention, consciousness, and data display," *American Statistical Association, Statistical Graphics Section*, 2006.
- [9] D. A. Szafir, "Modeling color difference for visualization design," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 392–401, 2018.
- [10] B. Preim, A. Baer, D. Cunningham, T. Isenberg, and T. Ropinski, "A survey of perceptually motivated 3d visualization of medical image data," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 501–525.
- [11] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *IEEE Symposium on Information Visualization (InfoVis)*, 2005, pp. 111–117.
- [12] G. J. Quadri and P. Rosen, "You can't publish replication studies (and how to do it anyway)," *arXiv preprint arXiv:1908.08893*, 2019.
- [13] R. Borgo, L. Micallef, B. Bach, F. McGee, and B. Lee, "Information visualization evaluation using crowdsourcing," *Computer Graphics Forum*, vol. 37, no. 3, pp. 573–595, 2018.
- [14] A. Diehl, A. Abdul-Rahman, M. El-Assady, B. Bach, D. A. Keim, and M. Chen, "Visguides: A forum for discussing visualization guidelines," in *EuroVis (Short Papers)*, 2018, pp. 61–65.
- [15] R. A. Rensink, "On the prospects for a science of visualization," in *Handbook of human centric visualization*. Springer, 2014.
- [16] M. A. Elliott, C. Nothelfer, C. Xiong, and D. A. Szafir, "A design space of vision science methods for visualization research," *IEEE Trans. on Visualization and Computer Graphics*, 2020.
- [17] M. Behrisch, M. Blumenschein, N. W. Kim, L. Shao, M. El-Assady, J. Fuchs, D. Seebacher, A. Diehl, U. Brandes, H. Pfister *et al.*, "Quality metrics for information visualization," *Computer Graphics Forum*, vol. 37, no. 3, pp. 625–662, 2018.
- [18] B. Saket, A. Srinivasan, E. Ragan, and A. Endert, "Evaluating interactive graphical encodings for data visualization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, 2018.
- [19] B. Saket, A. Endert, and C. Demiralp, "Task-based effectiveness of basic visualizations," *IEEE Trans. on Visualization and Computer Graphics*, 2018.
- [20] J. Talbot, V. Setlur, and A. Anand, "Four experiments on the perception of bar charts," *IEEE Trans. on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2152–2160, 2014.
- [21] L. Harrison, D. Skau, S. Franconeri, A. Lu, and R. Chang, "Influencing visual judgment through affective priming," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2013.
- [22] F. Yang, L. Harrison, R. Rensink, S. Franconeri, and R. Chang, "Correlation judgment and visualization features: A comparative study," *IEEE Trans. on Vis. and Computer Graphics*, 2018.
- [23] M. Brehmer, B. Lee, P. Isenberg, and E. K. Choe, "A comparative evaluation of animation and small multiples for trend visualization on mobile phones," *IEEE Trans. on Visualization and Computer Graphics*, 2019.
- [24] J. Choo, H. Lee, J. Kihm, and H. Park, "ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2010, pp. 27–34.
- [25] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager: Exploratory analysis via faceted browsing of visualization recommendations," *IEEE Trans. on Visualization and Computer Graphics*, no. 1, pp. 1–1, 2016.
- [26] B. Saket, H. Kim, E. Brown, and A. Endert, "Visualization by demonstration: An interaction paradigm for visual data exploration," *IEEE Trans. on Visualization and Computer Graphics*, no. 1, pp. 331–340, 2016.
- [27] L. Harrison, F. Yang, S. Franconeri, and R. Chang, "Ranking visualizations of correlation using weber's law," *IEEE Trans. on Visualization and Computer Graphics*, vol. 20, 2014.
- [28] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013.
- [29] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Transactions on Graphics (TOG)*, vol. 5, no. 2, pp. 110–141, 1986.
- [30] C. Demiralp, M. Bernstein, and J. Heer, "Learning perceptual kernels for visualization design," *IEEE Trans. on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1933–1942, 2014.
- [31] C. Gramazio, K. Schloss, and D. Laidlaw, "The relation between visualization size, grouping, and user performance," *IEEE Trans. on Visualization and Computer Graphics*, 2014.
- [32] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2009.
- [33] J. Heer and M. Bostock, "Crowdsourcing graphical perception: using mechanical turk to assess visualization design," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2010.
- [34] L. Neumann, K. Matkovic, and W. Purgathofer, "Perception based color image difference," *Computer Graphics Forum*, vol. 17, 1998.
- [35] K. Reda, P. Nalawade, and K. Ansah-Koi, "Graphical perception of continuous quantitative maps: the effects of spatial frequency and colormap design," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2018, p. 272.
- [36] M. Sedlmaier, A. Tatu, T. Munzner, and M. Tory, "A taxonomy of visual cluster separation factors," *Computer Graphics Forum*, vol. 31, no. 3pt4, pp. 1335–1344, 2012.
- [37] D. Chung, D. Archambault, R. Borgo, D. Edwards, R. Laramee, and M. Chen, "How ordered is it? on the perceptual orderability of visual channels," *Computer Graphics Forum*, vol. 35, no. 3, pp. 131–140, 2016.
- [38] V. Setlur and M. Stone, "A linguistic approach to categorical color assignment for data visualization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, 2016.
- [39] S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer, "Selecting semantically-resonant colors for data visualization," *Computer Graphics Forum*, vol. 32, no. 3pt4, pp. 401–410, 2013.
- [40] Y. Kim and J. Heer, "Assessing effects of task and data distribution on the effectiveness of visual encodings," *Computer Graphics Forum*, vol. 37, no. 3, pp. 157–167, 2018.
- [41] T. L. Turton, C. Ware, F. Samsel, and D. H. Rogers, "A crowdsourced approach to colormap assessment," in *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, 2017.
- [42] S. Smart, K. Wu, and D. A. Szafir, "Color crafting: Automating the construction of designer quality color ramps," *IEEE Trans. on Visualization and Computer Graphics*, 2019.
- [43] J. Li, J. van Wijk, and J.-B. Martens, "A model of symbol lightness discrimination in sparse scatterplots," in *IEEE Pacific Visualization Symposium (PacificVis)*, 2010, pp. 105–112.
- [44] S. L'Yi, Y. Chang, D. Shin, and J. Seo, "Toward understanding representation methods in visualization recommendations through scatterplot construction tasks," *Computer Graphics Forum*, vol. 38, no. 3, pp. 201–211, 2019.

[45] E. Dimara, A. Bezerianos, and P. Dragicevic, "Conceptual and methodological issues in evaluating multidimensional visualizations for decision support," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 749–759, 2018.

[46] D. Skau and R. Kosara, "Readability and precision in pictorial bar charts," in *Eurographics/IEEE VGTC Conference on Visualization (EuroVis): Short Papers*, 2017, pp. 91–95.

[47] C. Nothelfer and S. Franconeri, "Measures of the benefit of direct encoding of data deltas for data pair relation perception," *IEEE Trans. on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 311–320, 2019.

[48] H. Siirtola, T. Laivo, T. Heimonen, and K.-J. Raiha, "Visual perception of parallel coordinate visualizations," in *International Conference Information Visualisation (IV)*, 2009, pp. 3–9.

[49] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister, "Guidelines for effective usage of text highlighting techniques," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 489–498, 2015.

[50] E. Alexander, C.-C. Chang, M. Shimabukuro, S. Franconeri, C. Collins, and M. Gleicher, "Perceptual biases in font size as a data encoding," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 8, pp. 2397–2410, 2018.

[51] B. Yost and C. North, "The perceptual scalability of visualization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 837–844, 2006.

[52] S. Lee, M. Sips, and H.-P. Seidel, "Perceptually driven visibility optimization for categorical data visualization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, 2013.

[53] L. Battle, R. J. Crouser, A. Nakashimana, A. Montoly, R. Chang, and M. Stonebraker, "The role of latency and task complexity in predicting visual search behavior," *IEEE Trans. on Visualization and Computer Graphics*, 2019.

[54] R. Beecham, J. Dykes, W. Meulemans, A. Slingsby, C. Turkay, and J. Wood, "Map lineups: effects of spatial structure on graphical inference," *IEEE Trans. on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 391–400, 2017.

[55] S. Nusrat, M. J. Alam, and S. Kobourov, "Evaluating cartogram effectiveness," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 2, pp. 1077–1090, 2018.

[56] S. Afzal, R. Maciejewski, Y. Jang, N. Elmquist, and D. Ebert, "Spatial text visualization using automatic typographic maps," *IEEE Trans. on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2556–2564, 2012.

[57] C. Felix, S. Franconeri, and E. Bertini, "Taking word clouds apart: An empirical investigation of the design space for keyword summaries," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 657–666, 2018.

[58] J. Schrammel, M. Leitner, and M. Tscheligi, "Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2009, pp. 2037–2040.

[59] A. Rivadeneira, D. Gruen, M. Muller, and D. Millen, "Getting our head in the clouds: toward evaluation studies of tagclouds," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2007, pp. 995–998.

[60] C. Xiong, C. Ceja, C. Ludwig, and S. Franconeri, "Biased average position estimates in line and bar graphs: Underestimation, overestimation, and perceptual pull," *IEEE Trans. on Visualization and Computer Graphics*, 2019.

[61] Y. Liu and J. Heer, "Somewhere over the rainbow: An empirical assessment of quantitative colormaps," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2018, p. 598.

[62] M. Correll, D. Albers, S. Franconeri, and M. Gleicher, "Comparing averages in time series data," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2012, pp. 1095–1104.

[63] J. Li, J. van Wijk, and J.-B. Martens, "A model of symbol lightness discrimination in sparse scatterplots," in *IEEE Pacific Visualization Symposium (PacificVis)*, 2010.

[64] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri, "Perception of average value in multiclass scatterplots," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, 2013.

[65] J. Hullman, "Why evaluating uncertainty visualization is error prone," in *Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 2016, pp. 143–151.

[66] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker, "Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making," *Computer Graphics Forum*, vol. 33, no. 3, pp. 371–380, 2014.

[67] M. Waldner, A. Diehl, D. Gracanin, R. Splechtna, C. Delrieux, and K. Matkovic, "A comparison of radial and linear charts for visualizing daily patterns," *IEEE Trans. on Visualization and Computer Graphics*, 2019.

[68] A. Srinivasan, M. Brehmer, B. Lee, and S. Drucker, "What's the difference?: Evaluating variations of multi-series bar charts for visual comparison tasks," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2018, p. 304.

[69] M. Siegrist, "The use or misuse of three-dimensional graphs to represent lower-dimensional data," *Behaviour and Information Technology*, vol. 15, no. 2, pp. 96–100, 1996.

[70] D. Skau, L. Harrison, and R. Kosara, "An evaluation of the impact of visual embellishments in bar charts," *Computer Graphics Forum*, vol. 34, no. 3, pp. 221–230, 2015.

[71] D. Skau and R. Kosara, "Arcs, angles, or areas: Individual data encodings in pie and donut charts," *Computer Graphics Forum*, vol. 35, no. 3, pp. 121–130, 2016.

[72] E. Bertini, N. Elmquist, and T. Wischgoll, "Judgment error in pie chart variations," in *Eurographics/IEEE VGTC Conference on Visualization (EuroVis): Short Papers*, 2016, pp. 91–95.

[73] B. Ondov, N. Jardine, N. Elmquist, and S. Franconeri, "Face to face: Evaluating visual comparison," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[74] N. Jardine, B. Ondov, N. Elmquist, and S. Franconeri, "The perceptual proxies of visual comparison," *IEEE Trans. on Visualization and Computer Graphics*, 2019.

[75] D. Albers, M. Correll, and M. Gleicher, "Task-driven evaluation of aggregation in time series visualization," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2014.

[76] M. Ghoniem, J.-D. Fekete, and P. Castagliola, "A comparison of the readability of graphs using node-link and matrix-based representations," in *IEEE Symposium on Information Visualization (InfoVis)*, 2004, pp. 17–24.

[77] S. Few, "Journey to zvinka: The making of a new chart," *Visual Business Intelligence*, 2017.

[78] P. Mylavarapu, A. Yalcin, X. Gregg, and N. Elmquist, "Ranked-list visualization: A graphical perception study," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2019, p. 192.

[79] R. Rosenbaum, J. Zhi, and B. Hamann, "Progressive parallel coordinates," in *IEEE Pacific Visualization Symposium (PacificVis)*, 2012, pp. 25–32.

[80] M. Halvey and M. Keane, "An assessment of tag presentation techniques," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 1313–1314.

[81] J. Li, J.-B. Martens, and J. van Wijk, "A model of symbol size discrimination in scatterplots," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2010, pp. 2553–2562.

[82] C. Yau, M. Karimzadeh, C. Surakitanbarnharn, N. Elmquist, and D. Ebert, "Bridging the data analysis communication gap utilizing a three-component summarized line graph," *Computer Graphics Forum*, vol. 38, no. 3, pp. 375–386, 2019.

[83] M. Correll, M. Li, G. Kindlmann, and C. Scheidegger, "Looks good to me: Visualizations as sanity checks," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[84] L. Wilkinson, "Visualizing big data outliers through distributed aggregation," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 256–266, 2017.

[85] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri, "Four types of ensemble coding in data visualizations," *Journal of Vision*, vol. 16, no. 5, pp. 11–11, 2016.

[86] H. Rosling, "GapMinder Foundation," vol. 91, 2009.

[87] C. Ware and M. Plumlee, "Designing a better weather display," *Information Visualization*, vol. 12, no. 3–4, pp. 221–239, 2013.

[88] J. O'Callaghan, "Human perception of homogeneous dot patterns," *Perception*, vol. 3, no. 1, pp. 33–45, 1974.

[89] L. Wilkinson, A. Anand, and R. Grossman, "Graph-theoretic scagnostics," in *IEEE Symposium on Information Visualization (InfoVis)*, IEEE, Oct 2005, pp. 157–164.

[90] Y.-H. Chan, C. Correa, and K.-L. Ma, "The generalized sensitivity scatterplot," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 10, pp. 1768–1781, 2013.

[91] Y. Wang, F. Han, L. Zhu, O. Deussen, and B. Chen, "Line graph or scatter plot? automatic selection of methods for visualizing trends in time series," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 2, pp. 1141–1154, 2018.

[92] R. G. Raidou, M. Eisemann, M. Breeuwer, E. Eisemann, and A. Vilanova, "Orientation-enhanced parallel coordinate plots," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 589–598, 2015.

[93] M. Correll and J. Heer, "Regression by eye: Estimating trends in bivariate visualizations," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2017, pp. 1387–1396.

[94] A. Sarikaya, M. Gleicher, and D. A. Szafir, "Design factors for summary visualization in visual analytics," *Computer Graphics Forum*, vol. 37, no. 3, pp. 145–156, 2018.

[95] A. Sarikaya and M. Gleicher, "Scatterplots: Tasks, data, and designs," *IEEE Trans. on Visualization and Computer Graphics*, no. 1, pp. 402–412, 2018.

[96] S. Smart and D. A. Szafir, "Measuring the separability of shape, size, and color in scatterplots," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2019, p. 669.

[97] Y. Sadahiro, "Cluster perception in the distribution of point objects," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 34, no. 1, pp. 49–62, 1997.

[98] L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf, "Towards perceptual optimization of the visual design of scatterplots," *IEEE Trans. on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1588–1599, 2017.

[99] Y. Wang, X. Chen, T. Ge, C. Bao, M. Sedlmair, C.-W. Fu, O. Deussen, and B. Chen, "Optimizing color assignment for perception of class separability in multiclass scatterplots," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[100] J. Matejka, F. Anderson, and G. Fitzmaurice, "Dynamic opacity optimization for scatter plots," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2015, pp. 2707–2710.

[101] G. J. Quadri and P. Rosen, "Modeling the influence of visual density on cluster perception in scatterplots using topology," *IEEE Trans. on Visualization and Computer Graphics*, 2021.

[102] S. Few and P. Edge, "Solutions to the problem of over-plotting in graphs," *Visual Business Intelligence Newsletter*, 2008.

[103] A. C. Valdez, M. Ziefle, and M. Sedlmair, "Priming and anchoring effects in visualization," *IEEE Trans. on Visualization and Computer Graphics*, no. 1, pp. 584–594, 2018.

[104] R. Allen and M. Rubin, "Chromostereopsis," *Survey of Ophthalmology*, vol. 26, no. 1, pp. 22–27, 1981.

[105] W. C. Eells, "The relative merits of circles and bars for representing component parts," *Journal of the American Statistical Association*, vol. 21, no. 154, pp. 119–132, 1926.

[106] I. Spence and S. Lewandowsky, "Displaying proportions and percentages," *Applied Cognitive Psychology*, vol. 5, no. 1, 1991.

[107] M. Sedlmair and M. Aupetit, "Data-driven evaluation of visual quality measures," *Computer Graphics Forum*, vol. 34, 2015.

[108] M. Aupetit and M. Sedlmair, "Sepme: 2002 new visual separation measures," in *IEEE Pacific Visualization Symposium (PacificVis)*, 2016.

[109] Y. Ma, A. Tung, W. Wang, X. Gao, Z. Pan, and W. Chen, "Scatternet: A deep subjective similarity model for visual analysis of scatterplots," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[110] T. N. Dang and L. Wilkinson, "Transforming scagnostics to reveal hidden features," *IEEE Trans. on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1624–1632, 2014.

[111] J. Matute, A. Telea, and L. Linsen, "Skeleton-based scagnostics," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 542–552, 2017.

[112] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini, "Towards understanding human similarity perception in the analysis of large sets of scatter plots," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2016, pp. 3659–3669.

[113] M. Abbas, M. Aupetit, M. Sedlmair, and H. Bensmail, "Clustme: A visual quality measure for ranking monochrome scatterplots based on cluster patterns," *Computer Graphics Forum*, vol. 38, no. 3, pp. 225–236, 2019.

[114] H. Nguyen and P. Rosen, "Dspcp: A data scalable approach for identifying relationships in parallel coordinates," *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, no. 3, pp. 1301–1315, 2018.

[115] B. C. Kwon and B. Lee, "A comparative evaluation on online learning approaches using parallel coordinate visualization," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2016, pp. 993–997.

[116] D. Holten and J. J. Van Wijk, "Evaluation of cluster identification performance for different pcp variants," in *Computer Graphics Forum*, vol. 29, no. 3. Wiley Online Library, 2010, pp. 793–802.

[117] F. van Ham and B. Rogowitz, "Perceptual organization in user-generated graph layouts," *IEEE Trans. on Visualization and Computer Graphics*, no. 6, pp. 1333–1339, 2008.

[118] R. Rensink and G. Baldridge, "The perception of correlation in scatterplots," *Computer Graphics Forum*, vol. 29, no. 3, pp. 1203–1210, 2010.

[119] W. Cleveland, P. Diaconis, and R. McGill, "Variables on scatterplots look more highly correlated when the scales are increased," *Science*, vol. 216, no. 4550, pp. 1138–1141, 1982.

[120] J. Li, J.-B. Martens, and J. van Wijk, "Judging correlation from scatterplots and parallel coordinate plots," *Information Visualization*, vol. 9, no. 1, pp. 13–30, 2010.

[121] R. Chang, F. Yang, and M. Procopio, "From vision science to data science: Applying perception to problems in big data," *Electronic Imaging*, vol. 2016, no. 16, pp. 1–7, 2016.

[122] M. Kay and J. Heer, "Beyond weber's law: A second look at ranking visualizations of correlation," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, 2016.

[123] J. Zacks and B. Tversky, "Bars and lines: A study of graphic communication," *Memory and Cognition*, vol. 27, 1999.

[124] P. Bobko and R. Karren, "The perception of pearson product moment correlations from bivariate scatterplots," *Personnel Psychology*, vol. 32, no. 2, pp. 313–325, 1979.

[125] T. Lauer and G. Post, "Density in scatterplots and the estimation of correlation," *Behaviour and Information Technology*, vol. 8, 1989.

[126] J. Meyer, M. Taieb, and I. Flascher, "Correlation estimates as perceptual judgments," *J. of Experimental Psychology: Applied*, 1997.

[127] I. Pollack, "Identification of visual cor relational scatterplots," *Journal of Experimental Psychology*, vol. 59, no. 6, p. 351, 1960.

[128] H. Nguyen and P. Rosen, "Correlation coordinate plots: Efficient layouts for correlation tasks," in *International Joint Conference on Computer Vision, Imaging and Computer Graphics*, 2016, pp. 264–286.

[129] V. Sher, K. Bemis, I. Liccardi, and M. Chen, "An empirical study on the reliability of perceiving correlation indices using scatterplots," *Computer Graphics Forum*, vol. 36, no. 3, pp. 61–72, 2017.

[130] L. Best, A. Hunter, and B. Stewart, "Perceiving relationships: A physiological examination of the perception of scatterplots," in *International Conference on Theory and Application of Diagrams*, 2006.

[131] V. Peña-Araya, E. Pietriga, and A. Bezerianos, "A comparison of visualizations for identifying correlation over space and time," *IEEE Trans. on Visualization and Computer Graphics*, 2019.

[132] E. Zgraggen, Z. Zhao, R. Zeleznik, and T. Kraska, "Investigating the effect of the multiple comparisons problem in visual analysis," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2018.

[133] T. Wun, J. Payne, S. Huron, and S. Carpendale, "Comparing bar chart authoring with microsoft excel and tangible tiles," *Computer Graphics Forum*, vol. 35, no. 3, pp. 111–120, 2016.

[134] W. Javed, B. McDonnel, and N. Elmquist, "Graphical perception of multiple time series," *IEEE Trans. on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 927–934, 2010.

[135] H. Reijner, "The development of the horizon graph," *White-paper*, 2008.

[136] C. Perin, F. Vernier, and J.-D. Fekete, "Interactive horizon graphs: Improving the compact visualization of multiple time series," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2013.

[137] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos, "Comparing similarity perception in time series visualizations," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[138] E. Tufte, *The visual display of quantitative information*. Graphics press Cheshire, CT, 2001, vol. 2.

[139] C. Chang, B. Bach, T. Dwyer, and K. Marriott, "Evaluating perceptually complementary views for network exploration tasks," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2017.

[140] H. Guo, J. Huang, and D. Laidlaw, "Representing uncertainty in graph edges: An evaluation of paired visual variables," *IEEE Trans. on Visualization and Computer Graphics*, 2015.

[141] U. Soni, Y. Lu, B. Hansen, H. Purchase, S. Kobourov, and R. Maciejewski, "The perception of graph properties in graph layouts," *Computer Graphics Forum*, vol. 37, no. 3, 2018.

[142] R. Kosara, "An empire built on sand: Reexamining what we think we know about visualization," in *BELIV*, 2016.

[143] K. Hornbæk, S. Sander, J. A. Bargas-Avila, and J. Grue Simonsen, "Is once enough? on the extent and content of replications in

human-computer interaction," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2014, pp. 3523–3532.

[144] C. Wacharamanotham, L. Eisenring, S. Haroz, and F. Echtler, "Transparency of chi research artifacts: Results of a self-reported survey," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2020.

[145] J. M. Loomis, J. J. Blascovich, and A. C. Beall, "Immersive virtual environment technology as a basic research tool in psychology," *Behavior research methods, instruments, & computers*, vol. 31, no. 4, pp. 557–564, 1999.

[146] S. Carpendale, "Evaluating information visualizations," in *Information visualization*. Springer, 2008, pp. 19–45.

[147] S. Gelfand, *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2017.

[148] G. Gescheider, *Psychophysics: the fundamentals*. Psychology Press, 2013.

[149] E. H. Weber, *De pulsu, respiratione, auditu et tactu: annotationes anatomicae et physiologicae, auctore: prostat apud CF Koehler*, 1834.

[150] ———, *EH Weber on the tactile senses*. Psychology Press, 1996.

[151] E. Weber, "Weber's law of just noticeable differences," USD Internet Sensation & Perception Laboratory, 1834.

[152] S. Coren, L. Ward, and J. Enns, "Sensation and perception," 1999.

[153] A. C. Valdez, M. Ziefle, and M. Sedlmaier, "A framework for studying biases in visualization research," in *DECISIVe 2017*, 2017.

[154] M. Morgan, G. Hole, and A. Glennerster, "Biases and sensitivities in geometrical illusions," *Vision Research*, vol. 30, 1990.

[155] T. Seizova-Cajic and B. Gillam, "Biases in judgments of separation and orientation of elements belonging to different clusters," *Vision Research*, vol. 46, no. 16, pp. 2525–2534, 2006.

[156] L.-Q. Ma, K. Xu, T.-T. Wong, B.-Y. Jiang, and S.-M. Hu, "Change blindness images," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 11, pp. 1808–1819, 2013.

[157] W.-C. Wu, C. Basdogan, and M. Srinivasan, "Visual, haptic, and bimodal perception of size and stiffness in virtual environments," *ASME Dynamic Systems and Control Division*, 1999.

[158] K. Zibrek, E. Kokkinara, and R. McDonnell, "The effect of realistic appearance of virtual characters in immersive environments—does the character's personality play a role?" *IEEE Trans. on Visualization and Computer Graphics*, vol. 24, 2018.

[159] H. Song and D. A. Szafir, "Where's my data? evaluating visualizations with missing data," *IEEE Trans. on Visualization and Computer Graphics*, 2018.

[160] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. Keim, "The role of uncertainty, awareness, and trust in visual analytics," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 240–249, 2016.

[161] A. V. Pandey, K. Rall, M. Satterthwaite, O. Nov, and E. Bertini, "How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques," in *ACM SIGCHI Conf. on Human Factors in Comp. Systems*, 2015, pp. 1469–1478.

[162] G. Ellis, "So, what are cognitive biases?" in *Cognitive Biases in Visualizations*. Springer, 2018, pp. 1–10.

[163] E. Dimara, A. Bezerianos, and P. Dragicevic, "The attraction effect in information visualization," *IEEE Trans. on Visualization and Computer Graphics*, vol. 23, no. 1, 2017.

[164] Y. P. Sue O'Curry and R. Pitts, "The attraction effect and political choice in two elections," *Journal of Consumer Psychology*, vol. 4, no. 1, pp. 85–101, 1995.

[165] D. Meyer and R. Schvaneveldt, "Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations," *J. of Experimental Psychology*, vol. 90, 1971.

[166] L. Harrison, R. Chang, and A. Lu, "Exploring the impact of emotion on visual judgement," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2012, pp. 227–228.



Ghulam Jilani Quadri is a fourth-year Ph.D. student at the University of South Florida in the Department of Computer Science and Engineering under the supervision of Dr. Paul Rosen. Before joining the University of South Florida, Ghulam worked for Infosys Limited as System Engineer in Pune. Ghulam and team participated in IEEE VIS 2017 VAST Challenge and awarded Honorable mention. Ghulam received a 2021 Computing Innovation Fellow award. His current research interests are Information Visualization and HCI.



Paul Rosen is an Associate Professor at the University of South Florida. He received his Ph.D. from the Computer Science Department of Purdue University. His research interests include applying geometry- and topology-based approaches to problems in information visualization. Along with his collaborators, he has received awards for best paper at PacificVis 2016, IVAPP 2016, PacificVis 2014, and SIBGRAPI 2013. Dr. Rosen received a National Science Foundation CAREER Award in 2019.